

Result No.	Query			DB	ID	Description
	Score	Match	Length			
1	1232.5	37.5	623	3	P79076	P79076 coriolus ve
2	203	6.2	551	2	Q9RH54	Q9RH54 pantoea agg
3	200.5	6.1	573	16	Q9P190	Q9P190 campylobact
4	200	6.1	545	2	Q9LBR8	Q9LBR8 gluconobact
5	191.5	5.8	615	2	Q34214	Q34214 pectobacter
6	190.5	5.8	579	16	Q9AT76	Q9AT76 caulobacter
7	182	5.5	591	16	Q9I1K8	Q9I1K8 pseudomonas
8	180	5.5	553	2	Q9XCR0	Q9XCR0 pantoea cit
9	179.5	5.5	529	17	Q9HQ8R	Q9HQ8R halobacteri
10	176	5.4	722	16	Q9RZ26	Q9RZ26 deinococcus
11	158.5	4.8	748	10	Q9M0H4	Q9M0H4 arabidopsi
12	158.5	4.8	748	10	Q9ABP3	Q9ABP3 arabidopsi
13	155.5	4.7	499	16	Q988P2	Q988P2 rhizobium l
14	155.5	4.7	769	3	O74253	O74253 pycnopus
15	149	4.5	479	16	Q98D66	Q98D66 rhizobium l
16	141	4.3	523	16	Q98C76	Q98C76 rhizobium l

DE GLUCONATE DEHYDROGENASE.
GN Pseudomonas aeruginosa.
OS Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Pseudomonas
OX NCBI_TaxID=287;
RN [1]
RN SEQUENCE FROM N.A.
RP STRAIN=ATCC 15692 / PAOI;
RC MEDLINE=2043737; PubMed=10984043;
RX Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warren P.,
RA Hickey R.J., Brinkman F.S.T., Hufnagle W.O., Kowalik D.J., Lagrou M.,
RA Garber R.L., Golltry L., Tolentino E., Westbrock-Wadman S., Yuan Y.,
RA Brody L.A., Coulter S.H., Folger K.R., Kas A., Larbig K., Lim R.M.,
RA Smith K.A., Spencer D.N., Wong G.K.-S., Wu Z., Paulsen I.T.,
RA Reizer J., Sailer M.H., Hancock R.E.W., Lory S., Olson M.V.;
RT "Complete genome sequence of Pseudomonas aeruginosa PAOI, an
RT opportunistic pathogen.";
RT Nature 406:959-964(2000).
DR EMBL; AE004652; AAG05653.1;
KW Complete proteome.
SQ SEQUENCE 591 AA: 64716 MW; 9A7012B67ECEB9DE CRC64;

Query Match 5.5%; Score 182; DB 16; Length 591;
Best Local Similarity 20.6%; Pred. No. 4.9e-06;
Matches 138; Conservative 66; Mismatches 261; Indels 204; Gaps 23;

QY 36 VDVEFIAGSPIGATYAKLCVEAGLRVVMVEIGAADSFAVNAEGTAVPYVGHKNEI 95
DB 8 VDAVVGFWGTGAILAKELTEAGLNVALERGNRTYDGAYPNTL-----DEL 57
QY 96 EFQKIDRFVNVKIGALQOVSVFVRNQNVPTLDPGASAPPGSSAISNGKPHOREFNL 155
DB 58 TYNTRCKLFQNLKSTV-----SIRGINDTALPYRQL 90
QY 156 SAAVTRGVGGSTHTWCSTPRIHPMESLPG-----IGRKLNDPAEDDEKN 205
DB 91 SAEPLGQDGVGGAGLWSGVHFRIMPEELRLSHYEYRGKFIPEGMTIQDYGVSYELE 150
QY 206 ELYSEAEERLIGTSTKEFDESIRHTL-----VLSLQDAYKDRQIF 246
DB 151 PHDFAEKVFQTS-----GTATHTVQGVQVKGPNFAADRSDDPLPALRQVYS--AQLF 202
QY 247 R-----PLPLA-----CHRLKNAPPEYVEHSAENLFHSIY 276
DB 203 RKAEEELGHPYDLPAANASGPWNPYGVQMGPCNFCGFCGYACMYSKASPNL--NIL 260
QY 277 NDDKQKKLFTLTNHRCTFLATGGYEKKGA-----AEVRNLLATRNPSOLDYSIMAKV 332
DB 261 PALRQPLFELRANCVLKNVLDSDRGATGVYVDAQGREIVQ-----AKL 308
QY 333 YVLASCAIGNPOLYNSFGSGVQVTPRNDLSIPNLGRVITTEQPMACQIVLRQEFVDSR 392
DB 309 VIISAFQFNVRLLLSGI--GRPYDPRTEGV--VGKNFAQNMA-----TIKAFEDKDVH 361
QY 393 DDY-----GLPWWKEAQAQIAKNPTDALPIPRDPE 425
DB 362 TNPFGVTGGGVAVDDFNADNFDHGFLGVGSPMW-----VQAGSKPIGGLAVPQGPS 417
QY 426 PQVTTFTTEEHWHWT---QIHRDAFSYGAAGPEVDSRVIVDLRFEGATDPANNLLVFQN 482
DB 418 -----WGSQKQAVKDAYTH-----TVSMDAHGNSMTYRDNYLDLP 454
QY 483 DVODGYSMP--OPTFRY-----RPTASNVARKMADMCEVA-----SNLGGVLPSP 530
DB 455 TYKDATGQPLLRITFDWKDNETMGRYTERMRKTAEMNPKAISVSKNFGDHFNTFRV 514
QY 531 QFMDPGLALHAGTTRIGDKATTNADNSLWDFANLYVAGNCTIRTGFGENPLTSMC 590
DB 515 Q-----TTHLLGGAIMGSDPKTSVLNRYLOSVDVHNVFVVGASAFPGQGTGYNPTGLVAA 568
QY 591 HAITSARSI 599

Db 569 LAYWSAKAI 577

RESULT 8
Q9XCR0 PRELIMINARY; PRT; 553 AA.
ID Q9XCR0;
AC Q9XCR0;
DT 01-NOV-1999 (TREMBlrel. 12, Created)
DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE 2-KETO-GLUCONATE DEHYDROGENASE SUBUNIT.
GN KDBG.
OS Pantoea citrea.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Pantoea.
OX NCBI_TaxID=53336;
RN [1]
RN SEQUENCE FROM N.A.
RP STRAIN=1056R;
RC MEDLINE=20200361; PubMed=10735866;
RX Pujol C.J., Kado C.I.;
RA "Genetic and Biochemical characterization of the pathway in Pantoea
RT citrea leading to pink disease of pineapple.";
RL J. Bacteriol. 182:2230-2237(2000).
DR EMBL; AF131202; AAD44706.1;
DR InterPro; IPR000205; NAD_binding.
SQ SEQUENCE 553 AA: 59939 MW; F2A60424416B8C3F CRC64;

Query Match 5.5%; Score 180; DB 2; Length 553;
Best Local Similarity 20.6%; Pred. No. 6.4e-06;
Matches 134; Conservative 69; Mismatches 260; Indels 186; Gaps 24;

QY 32 GNDGVDFVFIAGSPIGATYAKLCVEAGLRVVMVEIGAADSFAVNAEGTAVPYVGYHK 91
DB 12 GSAADIVIVGSGIVGGIADRLVSGQSVLLILEAGLRIS--RAQAVENRNRMPFA----- 65
QY 92 KNEIEFQKIDRFVNVKIGALQOVSVFVRNQNVPTLDPGASAPPGSSAISNGKPHORE 151
DB 66 -----NRAGSDFQGLYPQ-----SPLAPAPLYFPNNYV--NVTGPSAGS 103
QY 152 FENLSAEAVTRGVGGSTHTWCSTPRIHP-----MESLPGIGRP--KLSNDPAEDDEKNEL 207
DB 104 FQ-----QGYLRTVGGTTHWAASCRWHHPSPDFVMSKSYGVGRDWPISYDEME---PW--- 153
QY 208 YSEAEERLIGT-----STKEFDESIRHTLVLSLQDAY-----KDRQR 244
DB 154 YCEAEYFICVAGSPDSQSPSRSPSRPMDVPMVFAHGDTPASVNVPHGNLVPIPOGR 213
QY 245 IFRP---LPLAC-----HRLKNAPEYVEHSAENLFHSIYNDKQK 283
DB 214 STRPWEGRPVCCGNNQCPIGAMNGIIHIERAESKGAVALAESVYVYKIDTDD--- 269
QY 284 LFTLTNHRCTR---LALTGGYEKKIGAEVRNLLATRNPSOLDYSYIMAKVYVLASAI 340
DB 270 -----NNRVTAVHMLDNOGASHKATG-----KAFALACNGI 300
QY 341 GNPOILYNSFGSGVQVTPRNDLSIPN---LGR-----YITEQPMACQIVLRQ 385
DB 301 ETPRLQLQ-----ANKNPTGIANSMDVMGRNMDHSGFHCFLTEEPVWLGKGPQAS 354
QY 386 EFDVSDVRDDPYGLPWWKEAQAQIAKNPTDALPIPRDPEQVTPFTTEEHWHWTQIHRD 445
DB 355 SCWVGRDGAFRSEY---SANKMILNLSRVVPAT-----KQ 388
QY 446 AFSYGAVGPEVDSRYI-----VDLRWFGATDPEANLLLVFONDVQDGYSMPOPTFRYRP 499
DB 389 ALAKGLGVKALDEIRYSIHGVLDLSLEPLDPDENRLTSLKTRKDPHGLACPDHIDV 448
QY 500 STASNVARKMADMCEVANSNGLGGLPTSPPOFMDPGLAL-----HLACTTRIGDKATTV 555
DB 449 GDYVRKGATAAHEQLQHIGSLFNG-----KEFNITALLNANHIMGGTIMGSAKDVA 501

QY 187 GIGRPKLSN-----DPAEDDKENNELYSEAEIRLIGTSTKEFDESIRHTLVLSLQDA 238
 Db 294 ---RORWASEHGLSDVADPGYD-----RHIDAVLERM-GVSEQCSDHNGPHQ-----RLVEGA 342
 QY 239 YKDRQIRFRPLPLACHRLKNAPEVVEHSAENLFHSIYND-----KOKKLEF----- 286
 Db 343 DKLGITYFKA-----ALNLSPEH---YDADKAGHAGFGDQAKQGTINTFLKDAFEAG 393
 QY 287 ---LLTNHRCRLALTGYEKKIGAAEVRNLLATRNPSQLDSYIMAKVYVLASGAINPQ 344
 Db 394 ARILVGTFRQARVLVEDG-----RAAGVSAVTTMGDETROIT--VRAPQVVVACGALETPA 446
 QY 345 ILVNSGFSGLQVTRNDSLPNLRGRYTEQPMFAFCQIVLRQEFVDSVRDDPYG---LPPW 401
 Db 447 LLRSGLGG-----PAGRYLRLHAPGLVAGI-----YGEDORAWW 482
 QY 402 KEVAQHAIAKNPTDALPIPRDPE-----POVTT---PFT---EER-PHHTQI 442
 Db 483 GP-----POSGILKQFADHENGHGFIEGVQVGPALMASSGLPWTGGEAHRDLMSKF 533
 QY 443 HRDAFSYGAVPEVDSRIYD-----LRWEGATDP-EANLLVFQNDVQDGYSM-----P 491
 Db 534 HRMAFVSIYQDRHGQVTVDDGNAVHTYALTDLDARN---FRRGVTESIRLHEAAGA 590
 QY 492 QPTFRYRPSTASNVARKMMADMCEVAS---NLGGVLTSPPOFMDPGLALHLAGTTRIG 548
 Db 591 EEIVALAPGVAPWRGDDLEAFTGQVAQVPLGAGGQTVFS-----AHQMGSGARMG 640
 QY 549 FDKATTVADNNSLVMDFANLYVAGNGTIRTFGENPTLTSMCHAIKSARSIIINTLKGTTD 608
 Db 641 SDPQTSVADPDGQLHDVPGVWIGDTSAPFTCSGVNPMVSMALASRTAEKLLAAMEGADG 700
 QY 609 GKNTG 613
 Db 701 GTCSG 705
 RESULT 11
 Q94BP3 PRELIMINARY; PRT; 748 AA.
 ID Q94BP3
 AC Q94BP3
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
 DE HYPOTHETICAL 82.0 KDA PROTEIN.
 GN AT4G28570.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Lennard N., Quail N., Harris B., Rajandream M.A., Bartell B.G.,
 RA Mewes H.W., Lemcke K., Mayer K.F.X.;
 RA Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 RL [2]
 RN SEQUENCE FROM N.A.
 RP EU Arabidopsis sequencing project;
 RA Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AL161573; CAB81445.1;
 DR HSP; P22637; 3COX.
 DR InterPro; IPR000205; NAD_binding.
 KW Hypothetical protein.
 SQ SEQUENCE 748 AA; 81959 MW; 13CD183F5940DAC3 CRC64;

Query Match 4.8%; Score 158.5; DB 10; Length 748;
 Best Local Similarity 19.4%; Pred. No. 0.00058;
 Matches 126; Conservative 90; Mismatches 215; Indels 217; Gaps 28;
 QY 37 DVFIAGSGPIGATYAKLCVCEAGLRVVMVEIGAADSFAVNAEBGTAVPYVPGVHKKEITE 96

Db 239 DAVVVGSGGGVAAANLAKAGLKVLEKG---NYTAHDYSGLEVSPMLEYKGG--- 292
 QY 97 FOKDIDRFVNVIRKALQOVSVYVRNQNVPTLDPGAWSPAGSSAISNGKPNHREFENLS 156
 Db 293 -----GLLTIV-----DGK-----FMLLA 306
 QY 157 AEAVTRGVGCMSTHTWCTSTPRIHPMESLPGIGRPKLSNDPAEDDKENNELYSEAEIRLIG 216
 Db 307 GSAVG---GGTAVNWSAS---IRTPDHVL-----QEW-----SEGSKIKF 340
 QY 217 TSKKEF---DE-SIRHTLVLSLQDAYKDRQIRFRPLPLACHRL-----KNAPD--- 261
 Db 341 FGSQEQSANDVETIRIGVTERCVKHGFQ-----QVLRKGCERLGLQVESVRNPSPDH 395
 QY 262 -----YVEHSAENLFHSIYNDKOKKLEFLLTNHRCRLAL-----TGGYEK--- 306
 Db 396 YCGLCGYGCRAKNGTDQTLVDVAVENGAVILTGIAERFVLDNTSSSNERKKRCVGV 455
 QY 307 GAAYVRNLLATRNPSQLDSYIMAKVYVLASGAINPQILYNSGSLQVTPRNDLSLPN 366
 Db 456 FASSVGKGKGGK-----FIIEARVTVSSAGSLTLPMLSSGLKN-----PN 497
 QY 367 LGR-----YITEQPMFAFCQIVLRQEFVDSVR---DDPYGLPMMKEAVAQHAIAKN 412
 Db 498 IGRNLKLPVLTWGTGYPPEKDSFGKMEGGIITSVHHMD-----TESGCKAILEN 550
 QY 413 PTDALPIPRDPEPQVTPPTFTEHPHWTQIHRDAFSYGAVPEVDSRVIVDLRWFGATDP 472
 Db 551 PLIG-PASYAGLSPWVS-----GPDLKERMIIK---YGR-- 580
 QY 473 EANNLLVFQNDVQDGYSM---PQPTFRYRPSTASNVARKMMADMCEVAS---NLGGY--- 524
 Db 581 --AHLFALVRLDGLSGEYVMENEVYTRTKKDRNLRAQLRQALRVSAAGAIVEGTYSRD 638
 QY 525 -----LPTSPPOFMDPGLAL-----HLACTTRIGDPKATTVADNN 559
 Db 639 GQRMKCEAITKEAMEEFLDEVDVAVGGVGTGKGEVWTTFSAHQMGSCRGVTAEGALDEN 698
 QY 560 SLVMDFANLYVAGNGTIRTFGENPTLTSMCHAIKSARSIIINTLKGTT 607
 Db 699 GESWEAEGFVCDGSLPSAVGVNPMITIIQSTAVCISSKIVDSIQNTK 746
 RESULT 12
 Q94BP3 PRELIMINARY; PRT; 748 AA.
 ID Q94BP3
 AC Q94BP3
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
 DE HYPOTHETICAL 82.0 KDA PROTEIN.
 GN AT4G28570.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Yamada K., Liu S.X., Sakano H., Pham P.K., Banh J., Chung M.K.,
 RA Dale J.M., Gibson H.A., Goldsmith A.D., Jiang P.X., Lee J.M.,
 RA Quach H.L., Tang C., Toriumi M., Yu G., Bowser L., Carlinici P.,
 RA Chen H., Cheuk R., Hayashizaki Y., Ishida J., Jones T., Kaniya A.,
 RA Karlin-Neumann G., Kawai J., Kim C., Koesema E., Lam B., Lin J.,
 RA Meyers M.C., Miranda M., Narusaka M., Nguyen M., Palm C.J.,
 RA Sakurai T., Satou M., Seki M., Shinn P., Southwick A., Tracy S.E.,
 RA Shinozaki K., Davis R.W., Ecker J.R., Theologis A.;
 RT "Full length cDNA of gene At4g28570 (GI:7269712).";
 RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AY039977; AAK64154.1;
 KW Hypothetical protein.
 SQ SEQUENCE 748 AA; 81987 MW; E3EB613F5D56A0D1 CRC64;

RA Moukha S.M., Dumonceaux T.J., Record E., Archibald F.S.;
RT "Cloning and analysis of Pycnoporus cinnabarinus cellobiose
dehydrogenase";
RL Gene 234:23-33(1999).
DR EMBL: AF081574; AAC32197.1; -.
DR InterPro: IPR000172; GMC_oxred.
DR Pfam: PF00732; GMC_oxred; 1.
DR PROSITE: PS00624; GMC_OXRED_2; 1.
KW Oxidoreductase.
SQ SEQUENCE 769 AA; 81818 MW; 17C36159A3EB95DC CRC64;

Query Match 4.7%; Score 155.5; DB 3; Length 769;
Best Local Similarity 19.2%; Pred. No. 0.0011;
Matches 132; Conservative 68; Mismatches 219; Indels 269; Gaps 30;
QY 37 DVFIAGSGPIGATYAKLCVEAGLRVVMVEIGAADSFYAVNAEETGAVPVPCYHKKNIE 96
DB 232 DYIIVAGPGGIIADRLSEAGKVKILLERGPST-----AETG-GTYAPWAKSONLT 285
QY 97 FQKQIDRFVNVKIGALQOVVPRNQNTPLDPGAW-----SAPPGSSAISNGK- 145
DB 286 ED-----IPGLFESNET-----DPNPWCKDNTFPAGCLLGGSVNGAL 326
QY 146 --NPHQREFENLSAEVTRGVGGSTHTWCSTPRIHPWESLPGIGRPKLSNDPAEDDK 203
DB 327 YWLPDADFST-----ANGWPTNMGNHAPYTSKLQRLPSTDHPS----- 366
QY 204 WNELSEAEIRIGTSTKEFDESIRHTLVLSIODA-----YKDRQRIE----- 246
DB 367 -----ADGNRLQESA-----TVVQLQGGYQOITINDPDYK--HVFYSAF 410
QY 247 -----RPLPLACHRLKNAPEVVEHSAENLFHSTYNDKOKKFLTLTNHRCRRLALT 300
DB 411 DFINGORAGPVA-----TYFQTASARSNE--VYKD-----YTLVSQVLRNGSTIG 454
QY 301 GYEKKIGAAEVRNLLATNPSSQLDSYIMAKVYVLASGAIGNPQILYNSFGSG---LQVT 357
DB 455 --VRTNNTALGPDGIPLNPNRNV-----ILAAGSFGTPIRLFQSGIGPTDMIQTV 503
QY 358 PRNDSLILNLGRYITEOPMAFCQIVLROEFVDSVRDDPGLPWKKEAVAQIAKNPTDAL 417
DB 504 QSNPTAANL-----PPQSEWINLPVGGQVSNPS----- 533
QY 418 PIPTRDPEQVTPPTTEERP-----WHTQIHRDAFSYAGVGEVDSRVIVDLRW 466
DB 534 -----INLVFT--HPSIDAYENWADYNSNPRPADAQY-----LQSRGV--- 571
QY 467 FGATDPEANLLVP-----QNDVODGYSMPQPTFRPS----- 500
DB 572 FAGASPKLNFWRAYGSGDGKTRYAGTVRPGAASVNTSVAYNASQITITVYLSEGITSR 631
QY 501 -----TASNVR-----RKMDMCEVASNLGGVLTSPPEFMDPGLAL--- 539
DB 632 GRLGVDAALNKAITTPWLTDPVDKTIILQALHDVVSNNV-----PGLTLTP 681
QY 540 -----HLAGTTRIGFDKATTVADNNSLVWDFANLYVAGNGTIRT 578
DB 682 DHTOTLEQVAAAYDPATMCSNHWGAAGKIGSSPSTAVVDENTKVTNTDNLFIVDASIIPS 741
QY 579 GFGENPTLTSCHAIKSARSINILKGG 606
DB 742 LPVGNPHGALMSAEQAQAKIL-ALAGG 768
RESULT 15
Q98D66 PRELIMINARY; PRT; 479 AA.
AC Q98D66;
DT 01-OCT-2001 (Tremblrel. 18, Created)
DT 01-OCT-2001 (Tremblrel. 18, Last sequence update)
DT 01-OCT-2001 (Tremblrel. 18, Last annotation update)
DE DEHYDROGENASE SUBUNIT I.

GN MLL4837.
OS Rhizobium loti (Mesorhizobium loti).
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OX Phyllobacteriaceae; Mesorhizobium.
RN NCBI_taxID=381;
RP SEQUENCE FROM N.A.
RC STRAIN=MAFF303099;
RX MEDLINE=21082930; PubMed=11214968;
RA Kaneko T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,
RA Watanabe A., Idesawa K., Ishikawa A., Kawashima K., Kimura T.,
RA Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsuno A.,
RA Mochizuki I., Nakayama S., Nakazaki N., Shimpo S., Sugimoto M.,
RA Takeuchi C., Yamada M., Tabata S.;
RT "Complete genome structure of the nitrogen-fixing symbiotic bacterium
Mesorhizobium loti";
RL DNA Res. 7:331-338(2000).
RE EMBL: AP003005; BAB51405.1; -.
DR InterPro: IPR000205; NAD_binding.
DR InterPro: IPR003042; Rng_mnxygenase.
DR InterPro: IPR003975; Shal_channel.
DR PRINTS: PR00420; RNGMNOXGNASE.
DR PRINTS: PR01497; SHALCHANNEL.
KW Complete proteome.
SQ SEQUENCE 479 AA; 52874 MW; CB647738AF50800E CRC64;

Query Match 4.5%; Score 149; DB 16; Length 479;
Best Local Similarity 19.9%; Pred. No. 0.0017;
Matches 126; Conservative 59; Mismatches 198; Indels 250; Gaps 26;
QY 28 HETY--GNDGVDFVFIAGSGPIGATYAKLCVEAGLRVVMVEIGA-----ADSFYAVNAEE 79
DB 6 YEAYRAANFKPKVCILGSGPAGTTIARKLGAAGIPVWLEAGSRESDSQDY-----R 60
QY 80 GTAPVYVPGYHKKNIEFQKQIDRFVNVKIGALQOVVPRNQNTPLDPGWSAPPGSS 139
DB 61 GT-----KVGDFYFDLDIR--LRYNGGSSNHWAGWCRVLDLDFEPAW-AP----- 105
QY 140 AISNCKNPHQREFENLSAEVTRGVGGSTHTWCSTPRIHPWESLPGI-----GRPKLS 194
DB 106 -----DTGWPIRADIEPYLTVRDLLELDFDRPDV- 136
QY 195 NDPAEDDKENNELY--SEAEIRIGTSTKEFDESIRHTLVLSIODA YKDRQRIEPLPLAC 253
DB 137 --PISDDIRWQLIKSPAVRFAEKFADELKSGNIAVL----- 173
QY 254 HRLKNAPEVVEHSAENLFHSTYNDKOKKFLTLTNHRCRRLALTGGEYKKGAAEVRN 313
DB 174 -----NTYATELA--GDGKRVTKGLWS 194
QY 314 LLATRNPSQLDSYIMAKVYVLASGAIGNPQILYNSFGSLQVTPRNDSLIPN---LGRV 370
DB 195 -----NQDAGAFSADYFIVCTGGLNSRLLSN-----QRSNGGVVFNATALGRY 241
QY 371 ITEQP-----MAFCQIVLROEFVDSVRDDPYG----- 397
DB 242 WNEHPTFEGGNAILADYGEFEVDVAEAFSPPLAAMERLRIMNFGIRLIETPYAGVKHV 301
QY 398 -----LPWKEAVAQIAKN-----PTDALPTFRDPEPQVT-----TPFTSE 435
DB 302 IADLACTAPDMAEWAYQOLSONLRCAOALYVAWEQALSNQIELSKTDVDHAGVPRIEL 361
QY 436 HPWHTQIHRDAFSYAGVGEVDSRVIVD--LRWFGATDEANNLLVFONDYQGSNPQPT 494
DB 362 H-WKKS-----ELERTLVBEGLKLFGT-----LIEKDL----- 389
QY 495 FRYRPSSTASNVRAKRMADMCEVASNLGGVLTSPPEFMDPGLALHLAGTTRIGFDKAT 554
DB 390 -----GRVRLMDWIAN-----GEDYPTD-----EETAGHHHGGTRGIDMTSR 428
QY 555 VADNNSLVWDFANLYVAGNGTIRTGFGENPTLT 587

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 26, 2002, 11:07:23 ; Search time 13.49 Seconds
(without alignments)
1773.809 Million cell updates/sec

Title: US-09-856-327-2

Perfect score: 3284
Sequence: 1 MSLSTEQMLRDYPRSMQING.....IINTLKGDTGKNTGEHRNL 618

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

* Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	126.5	3.9	773	1 CDH_PHACH	Q01738 phanerochaete
2	116	3.5	514	1 GPD_MYCTU	O08407 mycobacteri
3	107.5	3.3	964	1 YIN0_YEAST	P40467 saccharomyc
4	103.5	3.2	612	1 AMYG_ASPOR	P36914 aspergillus
5	103.5	3.2	639	1 AMYG_ASPAK	P23176 aspergillus
6	102.5	3.1	1536	1 SIN3_YEAST	P22579 saccharomyc
7	101	3.1	1010	1 SCA4_RICPA	O9a175 rickettsia
8	100	3.0	663	1 ALOX_CANBO	Q00922 candida boi
9	100	3.0	1142	1 ENAM_FIG	O97939 sus scrofa
10	98.5	3.0	382	1 MTD_KLEPN	O9xbm6 klebsiella
11	98.5	3.0	556	1 BETA_ECOLI	P17444 escherichia
12	98	3.0	505	1 Y4NJ_RHISN	P55582 rhizobium s
13	98	3.0	925	1 W70T_HUMAN	P57737 homo sapien
14	98	3.0	1257	1 CCA4_BACTU	O45754 bacillus th
15	96.5	2.9	684	1 Y492_MYCTU	O11157 mycobacteri
16	96	2.9	291	1 U502_HSV11	P06485 herpes simp
17	95.5	2.9	382	1 MTD_ECOLI	P09424 escherichia
18	95.5	2.9	454	1 DLDH_RHOCA	P95596 rhodobacter
19	95.5	2.9	519	1 AMVH_SACFI	P26989 saccharomyc
20	95.5	2.9	708	1 HELS_SULSO	O974y9 sulfolobus
21	95.5	2.9	883	1 RCOL_EPT7	P00573 bacilliarop
22	95	2.9	487	1 ENGA_CHLFP	O92762 chlamydia p
23	95	2.9	500	1 YDAK_YEAST	P28817 schizosacch
24	94.5	2.9	855	1 GAF1_SCHPO	O10280 schizosacch
25	94.5	2.9	1211	1 BUN2_DROME	O24523 drosophila
26	94.5	2.9	2193	1 POLG_CXAL6	O65900 c genome po
27	94	2.9	606	1 PRIM_MYXIA	P50070 myxococcus
28	94	2.9	886	1 SM6B_MOUSE	O54951 mus musculu
29	93.5	2.8	406	1 YN05_YEAST	P53891 saccharomyc
30	93.5	2.8	594	1 CIK1_YEAST	O01649 saccharomyc
31	93.5	2.8	823	1 SCH9_YEAST	P11792 saccharomyc
32	93.5	2.8	903	1 VGLB_HSV1F	P06436 herpes simp
33	93.5	2.8	904	1 VGLB_HSV11	P10211 herpes simp

34 93.5 2.8 904 1 VGLB_HSV1P
35 93.5 2.8 1859 1 RP81_CAEEL
36 93 2.8 470 1 LEU2_AZOVI
37 93 2.8 552 1 CHOD_BREST
38 93 2.8 605 1 COX_ASPNG
39 93 2.8 737 1 AMY1_AEDAE
40 93 2.8 890 1 GLND_ECOLI
41 93 2.8 890 1 GLND_SALTY
42 93 2.8 953 1 YNM7_YEAST
43 93 2.8 955 1 VP2_BPV17
44 93 2.8 1012 1 UBAL_SCHPO
45 93 2.8 1117 1 CYT4_NEUCR

ALIGNMENTS

RESULT 1
CDH_PHACH STANDARD; PRT; 773 AA.
AC Q01738: 000047;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Cellobiose dehydrogenase precursor (EC 1.1.5.1) (CDH) (Cellobiose-
DE quinone oxidoreductase).
GN CDH-1 AND CDH-2.
OS Phanerochaete chrysosporium.
OC Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Homobasidiomycetes;
OC Aphyllophorales; Corticiales; Phanerochaete.
OX NCBI_TaxID=5306;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=OGC101;
RX MEDLINE=97077226; PubMed=8919793;
RA Li B., Nagalla S.R., Renganathan V.;
RT "Cloning of a cDNA encoding cellobiose dehydrogenase, a
RT hemoflavoenzyme from Phanerochaete chrysosporium.";
RL Appl. Environ. Microbiol. 62:1329-1335(1996).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=OGC101;
RX MEDLINE=97176414; PubMed=9023960;
RA Li B., Nagalla S.R., Renganathan V.;
RT "Cellobiose dehydrogenase from Phanerochaete chrysosporium is encoded
RT by two allelic variants.";
RL Appl. Environ. Microbiol. 63:796-799(1997).
RN [3]
RP X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS) OF 19-208.
RX MEDLINE=20139694; PubMed=10673428;
RA Hallberg B.M., Bergfors T., Boeckbro K., Pettersson G., Henriksson G.,
RA Divne C.;
RT "A new scaffold for binding haem in the cytochrome domain of the
RT extracellular flavocytochrome cellobiose dehydrogenase.";
RL Structure 8:79-88(2000).
CC -|- FUNCTION: DEGRADATES BOTH LIGNIN AND CELLULOSE. OXIDIZES CELLOBIOS
CC TO CELLOBIONOLACTONE.
CC -|- CATALYTIC ACTIVITY: Cellobiose + a quinone = cellobiono-1,5-
CC lactone + a phenol.
CC -|- COFACTOR: ONE FAD AND ONE HEME B.
CC -|- SUBCELLULAR LOCATION: Secreted.
CC -|- SIMILARITY: IN THE C-TERMINAL SECTION; BELONGS TO THE GMC
CC OXIDOREDUCTASES FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U46081; AAC49277.1; -

```
Query Match      3.5%; Score 116; DB 1; Length 514;
Best Local Similarity 22.8%; Pred. No. 0.34;
Matches 98; Conservative 48; Mismatches 163; Indels 120; Gaps 23;

QY 104 FNVVKGALQ--VSPVRNQNVTLPDQAWSPGSSAISNGKNPHOREFENLSAEAVTR 162
DB 77 FQGVYNAVOEHCRTFPFQONWDLAEG--FREVPQTFDDDAFAQLAETLEKLDAAE--R 132

QY 163 GVGGMSTHTWCSTPRIHPM--ESL--PGIGREKLSNDAEDDKENNELYSEARLIGTST 219
DB 133 GTGGNHAYLALPKSPFVVCQLHKSGLARPO-----GDRSVRVVIE----- 175

QY 220 KEFDESIRHTLVRLSLQADKYDRQIFRPLPLACHRL--KNAPEYVEMHSAENLFHSYN 277
DB 176 KPFCHDLASARELNKAVNAVPPEAEVR-----IDHVLCKETQNILARFANQLDFPIW 231

QY 278 ----DDQKKLFTLTNHRCTRLATGGYEKKIGAAE--VRN-----LATRNPSSQL 324
DB 232 AHYVDHVO-----ITMAEDIGLGRAGYDYGIAARDVIONHLMOLLALTAMEEPVSFH 285

QY 325 DSYIMA-KVYVLASGAIGNP-----QILYNSGFSG----- 353
DB 286 PAALQAEIKVLSTRLAEPLDOTTSGQYAGWGGEKVVGLLDEEGFAEDSTTETFAA 345

QY 354 --LQVTPRNDLIP-----NLGRYITEQPMACQIVLRQEFV--DSVRDDPYGLPWK 402
DB 346 ITLEVDTRRWAGVPEYLRGRLGRVTEIALVF---RRAPHLPDFATMTDELG----- 396

QY 403 EAVAQHIKNTDALPIFRDEPOVTPPTFEEHPWHQIHRDA---FSYGAV-----GPE 455
DB 397 -----TNAMVIRVO--PDEGTLRFSGVKPVTAMEVRDNNDFSYGSAFAEDSPE 444

QY 456 VDSRVIVDL 464
DB 445 AYERLILDV 453

RESULT 3
ID YINO YEAST STANDARD; PRT; 964 AA.
AC P40467;
* DT 01-FEB-1995 (Rel. 31, Created)
DE 16-FEB-1995 (Rel. 31, Last sequence update)
DE 16-OCT-2001 (Rel. 40, Last annotation update)
DE Putative 108.8 kDa transcriptional regulatory protein in FRK1-STH1
DE intergenic region.
DE Y11130W.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=288C / AB972;
RA Barrall B.G., Badcock K., Bankier A.T., Bowman S., Brown D.,
RA Churcher C.M., Connor K., Copsey T., Dear S., Devlin K., Fraser A.,
RA Gentles S., Hamlyn N., Horsnell T.S., Hunt S., Jagels K., Jones M.,
RA Louis E., Lye G., Moule S., Moule T., Odell C., Pearson D.,
RA Rajandream M.A., Riles L., Rowley N., Skelton J., Smith V.,
RA Walsh S.V., Whitehead S.;
RL Submitted (DEC-1994) to the EMBL/GenBank/DBJ databases.
CC -1- SUBCELLULAR LOCATION: Nuclear (Probable).
CC -1- SIMILARITY: CONTAINS 1 ZN(2)-CYS(6), FUNGAL-TYPE BINUCLEAR
CC CLUSTER DOMAIN.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
```

```
CC EMBL: Z38059; CAA86148.1;
DR PIR: S48404; S48404.
DR HSP: P08657; ICUD.
DR SGB: S0001392; Y11130W.
DR InterPro: IPR001138; ZN2_Cy6_fungal.
DR Pfam: PF00172; ZN_Cy6; 1.
DR PRINTS: PR00054; FUNGALZNCYS.
DR SMART: SM00066; GAL4; 1.
DR PROSITE: PS00463; ZN2_Cy6_FUNGAL_1; 1.
DR PROSITE: PS00463; ZN2_Cy6_FUNGAL_2; 1.
KW Hypothetical protein; Transcription regulation; DNA-binding;
KW Nuclear protein; Zinc; Metal-binding.
FT DOMAIN 21 47 ZN(2)-Cys(6), FUNGAL-TYPE.
FT DOMAIN 811 896 HIS-RICH.
SQ SEQUENCE 964 AA; 108780 MW; AD5ABE59E4B022CC CRC64;

Query Match      3.3%; Score 107.5; DB 1; Length 964;
Best Local Similarity 19.5%; Pred. No. 3.8;
Matches 113; Conservative 86; Mismatches 187; Indels 193; Gaps 30;

QY 33 NDGVDFYTAGSGPIGATYAK--LCVEAGLRVVMVEIGAADSFYVNAEGTAVPYV---- 86
DB 322 DEGYKYFAARKLIDITNARDLNSIQAILMLII-----FLQCSARLSTCYTYIGVAM 373

QY 87 ----PGYHKK-----NEIEFQK-----DIDRFVNVTKGALQOVSPVRNQP 125
DB 374 RSLRAGFHRKLSNPGSPGPIEIMKRKLFYTYIKUDYINAMGLUPRSISPDFFDTLP 433

QY 126 TLD-----PGAWSPGSAISNGKNPHOREF-----ENLSAEAVTRGVGMST 169
DB 434 -LDLSDENITEVAYLPENOHVSLSSTGIS---NEHTKFLILNEIISLYPIKKTNIIS 489

QY 170 HWTCT--PRIHPMESLPGIGRPKLSN--DPAEDDKENNELYSEARLIGTSTKEFDESI 226
DB 490 HETVTSLELKLRLNWLDSLPLKELIPNAENIDPE-----YERANRLHL-----SF 533

QY 227 RHTLVRLSLQDAYKDRQIFRPL-----PLACHRLKN-----APE 261
DB 534 LHVQII-----LYRFFIHYLSRNMNAENVPLCYRRARNSIAVARTVIKAKE 581

QY 262 YVE-----WHSAENLFHSIY-----NDDQKKLFTLTNHRCTRLALTGG 301
DB 582 MVSNNLLGTGYWYACVYTFYSVAGLLFYTHEAQLPKDSAREYDILKDAETGSRVLIQL 641

QY 302 YEKKGIAAEVRNLLATRNPSOLDYSYINAKVYVVLASGALGNFQILYN--SGFSGLOVTPR 359
DB 642 KDSSMAASRTYNLL-----NOIFEKLSNKTQLTA-----LHSSPSNESAFSLVTNN 687

QY 360 NDSLIPNLGRYITEQPMACQIVLRQEF-----VDSVRDDPYGLPWKKAQAQHIKNPT 414
DB 688 SSALAPHLGDSL-QPPVFFSSQDTKNSFLAKSESTND-----YAMANYLNTPI 737

QY 415 DALPIPRDPEPV---TTPFTEHPHWTQIHRDAFSYGAVGPEVDSRVIVDLRWFEGATD 471
DB 738 SENPLNEAQQDOVQSGQTNMSNE-----RDPNNFLSIIRLDNN----- 777

QY 472 PEANNLLVFQNDV---QDGYSMPQPT---FRYPSTASN 504
DB 778 -GQSNILDATDDVFIRNDG-DIPTNSAFDFSSSKSNASN 814.

RESULT 4
ID AMYG_ASPOR STANDARD; PRT; 612 AA.
AC P36914;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Glucosylase precursor (EC 3.2.1.3) (Glucan 1,4-alpha-glucosidase)
DE (1,4-alpha-D-glucan glucohydrolase).
GN GLAA.
```


RESULT 11

BETA_ECOLI STANDARD; PRT; 556 AA.

ID BETA_ECOLI

AC P17444; P77861;

DT 01-AUG-1990 (Rel. 15, Created)

DT 01-AUG-1990 (Rel. 15, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Choline dehydrogenase (EC 1.1.99.1) (CHD).

BETA OR B0311.

OS Escherichia coli.

OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;

OC Escherichia.

OX NCBI_TaxId=562;

[1]

RN SEQUENCE FROM N.A.

RP STRAIN=K12;

RC MEDLINE=92065800; PubMed=1956285;

RX Lamark T., Kaasen E., Eshoo M.W., Falkenberg P., McDougall J.,

RA Strom A.R.;

RT "DNA sequence and analysis of the bet genes encoding the

RT osmoregulatory choline-glycine betaine pathway of Escherichia coli.;"

RL Mol. Microbiol. 5:1049-1064(1991).

RN [2]

RP SEQUENCE FROM N.A.

RX MEDLINE=91348527; PubMed=1879697;

RA Boyd L.A., Adam L., Pelcher L.E., McHughen A., Hirji R., Selvaraj G.;

RT "Characterization of an Escherichia coli gene encoding betaine

RT aldehyde dehydrogenase (BADH): structural similarity to mammalian

RT ALDHs and a plant BADH.;"

RL Gene 103:45-52(1991).

RN [3]

RP SEQUENCE FROM N.A.

RC STRAIN=K12 / MG1655;

RX MEDLINE=97426617; PubMed=9278503;

RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,

RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,

RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,

RA Mau B., Shao Y.;

RT "The complete genome sequence of Escherichia coli K-12.;"

RL Science 277:1453-1474(1997).

RN [4]

RP SEQUENCE FROM N.A.

RA Duncan M., Allen E., Araujo R., Aparicio A.M., Chung E., Davis K.,

RA Federspiel N., Hyman R., Kaiman S., Komp C., Kurdi O., Lew H., Lin D.,

RA Namath A., Oefner P., Roberts D., Schramm S., Davis R.W.;

RL Submitted (OCT-1996) to the EMBL/GenBank/DBJ databases.

. Db 108 PGQOALPSAGVVLG--PEDLPVELVQFHTPSDGLVLSAAGTVVWVDAKQO---PLT 162
QY 184 SLPGIGRPKLSNPAEDDKENNELYSSEAERLICTSTKEFDESIRHTVLSLQDAYKDRQ 243
Db 163 EL-----AAHGLVQSAVWSRDGALVCT-----ACKDKQ 191
QY 244 -RIFRPLPLACHRLK-NAPEYVEWH-----SAENLPHSYNDKQKLL----- 284
Db 192 LRIFDP-----RTKPRASOSTOAHNSRSLRAWGTWEHLVSTGTGFNQMREREVLWD 245
QY 285 -----FTLLTNIRCT-----RLALFGGVEKKIGAAEVRNLLATRNPSQ--L 324
Db 246 RFPSSALASITLTSGLCLPLDPSGLLVLAGKERQLYCEYVVPQPALSPVTCVL 305
QY 325 DSIMAKVYVLAGAIGNPOILYNSGFLQVTPRNDSLIPNIGRYITEQPMACQIVLR 384
Db 306 ES-----VLRGAALVPROALAVMGCEVLRVLQSLDAITVPGYHVRKAVEFHEDL-- 356
QY 385 QEFVDSV---RDDPYGLPW--KEAVAOHIAKNP-----TDALPIPFDRPEQVIT 430
Db 357 --FPDTAGCVATDPHG--HWAGDNQOVQVSLNPACRPHSPSTSL-VPPAEPLPDTAQ 411
QY 431 PFTEHPWHTQIHRDAFSYGAVGPEVDSRVIVDLRFWFGATDPEANNLLVFQNDVQDGYM 490
Db 412 PAYMETP-----P-----VGDADASEGFSS 430
QY 491 PQTFRRYRSTASNV 505
Db 431 P-ESSLTSPSTPSSL 444

RESULT 14

CCAA_BACTU STANDARD; PRT; 1257 AA.
ID CCAA_BACTU
AC Q45754;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Pesticidal crystal protein cry12Aa (insecticidal delta-endotoxin
DE CryIIA(a)) (crystalline entomocidal protoxin) (142 kDa crystal
DE protein).
GN CRY12AA OR CRYIIA(A) OR CRYVB.
OS Bacillus thuringiensis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=1428;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 2-11.
RC STRAIN=NRRL B-18244 / PS33F2;
RA Narva K.E., Payne J.M., Schwab G.E., Hickie L.A., Galasan T.,
RA Sick A.J.;
RT 'Novel Bacillus thuringiensis microbes active against nematodes, and
RT genes encoding novel nematode-active toxins cloned from Bacillus
RT thuringi.';
RL Patent number EP0462721, 27-DEC-1991.
CC -1- FUNCTION: ENDOTOXIN WITH NEMATOCIDAL ACTIVITY.
CC -1- DEVELOPMENTAL STAGE: THE CRYSTAL PROTEIN IS PRODUCED DURING
CC SPOULATION AND IS ACCUMULATED BOTH AS AN INCLUSION AND AS PART
CC OF THE SPORE COAT.
CC -1- MISCELLANEOUS: TOXIC SEGMENT OF THE PROTEIN IS LOCATED IN THE
CC N-TERMINUS.
CC -1- SIMILARITY: BELONGS TO THE DELTA ENDOTOXIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/) or
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: L07027; AAA22355.1;
DR InterPro: IPR001178; Endotoxin.

DR Pfam: PF00555; endotoxin; 2.
KW Toxin; Sporulation.
SQ SEQUENCE 1257 AA; 142265 MW; 3D9888FFC6C0E3981 CRC64;

Query Match 3.0%; Score 98; DB 1; Length 1257;

Best Local Similarity 20.3%; Pred. No. 30;
Matches 132; Conservative 88; Mismatches 243; Indels 188; Gaps 36;

QY 15 SMOINGOIPKNAIHEVYGVNDGVDFIAGSPIGATYAKLCVEAGLRVVMVEIGAADSFYA 74
Db 300 SSPIYQVPKMNQNTSSIVPSDLFHYQGLVLEFSTRDNDGLAKIFT--GRNTFYK 357
QY 75 VNAEGTAVPYVPGYHKKNEIEFKDIDREVNVIKALQOVVVRNQNPT-LDPGAW 133
Db 358 -----SPNTHYHVDFSYNTSSGNSIRSGSNPDIPLNPIIISTCIRNSFYK 406
QY 134 APGSSAISNGKNPHQREFENLSAEAVTRGVGGMSHTWCTSPRIHPMESLPGIGPKL 193
Db 407 AIAGSSVLVNFKDGTO---GYAFAQAAPTGA-----WD-----HSFIES 442
QY 194 SNDAEDDKENNELYSSEAERLICTSTKEFDESIRHTLV-LRSLODAYKQRIFRPLPLA 252
Db 443 --DGAPEGHKLNIYITSP---GDTLDFINV--YILISTPTINELSTEIKGF---PAE 491
QY 253 CHRLLN-----APEYVEWHSAENLFHSIYNDKOKKLFLLTNHRCRTRALATGGYK 304
Db 492 KGYIKQGIKMYGKPEYINGAQPVL-----ENQOTLIFE-----HASKTAQYTI 538
QY 305 KIGAAEVRNL-----LATNPSSQLDSYIMAKVYVLAASGAIGNPOILYNSGSL 354
Db 539 RIRYASTQGTGYFRLDNDELQTLNPTSHNGYV-----TGNIGENDLYTIG--SY 588
QY 355 QVTPRNDSLIPNLGRYTEQPMACQI-----VL-ROEFV--DSVRDDPYGLPWKE 403
Db 589 TITEGHTL-----QIOHNDKNGWLDRIEFPKDSLQDSP----- 624
QY 404 AVAQHAKNPTDALPIFRDPEQV---TTPFTEHP--W-----HTQIHRDAFSYGAV 452
Db 625 -----QDSPEVHESTIIFDKSSPTIWSNNKHSYSHILEG--SYTSQ 665
QY 453 GPEVDSRVIVDLRFWFGATDPEANNLLVFQV-DYQDGYSMQPTFRYRPSASNVARKMM 511
Db 666 G-SYPHNLINL--FHPTDPNRNHTIHVNGDMNVYD-----KDSVADGLNFKIT 714
QY 512 ADMCEVASNLG-----GYLPTSPQFMDPGLALHLAG-TTRIGFKAT-----TVA 556
Db 715 AIPSDAWYSGTITSMHLENDNNFKITPKF---ELSNELNITTVQNALFASQAQDTLA 771
QY 557 DNNSLVMDPANLYVAGNGTIRTFG-ENPTLTSCHAIKARSIIINTLKG 606
Db 772 SNVSDYW-IEQVVMKVDALSDEVFGEKKALRLKLVNQAKRLSKIRNLLIG 821

RESULT 15

1492_MYCTU
ID Y492_MYCTU STANDARD; PRT; 684 AA.
AC Q11157;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical 57.3 kDa protein GMC-type oxidoreductase RV0492C.
GN RV0492C OR MT0512 OR MTCY2069.18C.
OS Mycobacterium tuberculosis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eigmeier K., Gas S., Barry C.E. III, Tekala F.,

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 26, 2002, 11:05:57 ; Search time 20.92 Seconds
(without alignments)
2838.585 Million cell updates/sec

Title: US-09-856-327-2
Perfect score: 3284
Sequence: 1 MSLSSTEQMLRDYPRSMQING.....IINTLKGGTGDGKNTGHRNL 618

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR-71.*

1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	200.5	6.1	573	2 E81385	probable oxidoredu
2	191.5	5.8	615	2 B38575	gluconate 2-dehydr
3	190.5	5.8	579	2 H87451	oxidoreductase, GM
4	187.5	5.7	619	2 A82003	dehydrogenase chai
5	182	5.5	591	2 H83362	gluconate dehydrog
6	179.5	5.5	529	2 A84260	hypothetical prote
7	176	5.4	722	2 B75608	GMC oxidoreductase
8	171.5	5.2	562	2 J7628	glucoside 3-dehydr
9	158.5	4.8	748	2 T10651	hypothetical prote
10	148.5	4.5	561	2 E98192	probable oxidoredu
11	148.5	4.5	561	2 A83094	oxidoreductase Atu
12	138.5	4.2	527	2 E98202	dehydrogenase chai
13	138.5	4.2	527	2 A83084	dehydrogenase Atu4
14	138.5	4.2	768	2 J65564	cellobiose oxidase
15	138	4.2	494	2 A83255	l-sorbose dehydrog
16	130.5	4.0	770	2 S60676	cellobiose oxidase
17	127.5	3.9	578	2 F70736	probable chob prot
18	124.5	3.8	578	2 D95400	probable oxidoredu
19	123	3.7	531	2 B83600	probable oxidoredu
20	121	3.7	1498	2 AF1082	B. subtilis Yuka p
21	117	3.6	736	2 A86171	hypothetical prote
22	116	3.5	502	2 T18562	hypothetical prote
23	116	3.5	514	2 B70917	probable zwf2 prot
24	116	3.5	599	2 T19711	hypothetical prote
25	114	3.5	1012	2 T00958	hypothetical prote
26	113.5	3.5	678	2 T05821	hypothetical prote
27	111.5	3.4	1042	2 T16169	hypothetical prote
28	110.5	3.4	589	2 T50698	probable mandeloni
29	110.5	3.4	5170	2 T15348	hypothetical prote

30 109.5 3.3 1125 2 T19193
31 108 3.3 1498 2 AG1439
32 107.5 3.3 557 2 F83181
33 107.5 3.3 567 2 AC0143
34 107.5 3.3 964 2 S48404
35 107 3.3 1018 2 T40253
36 106.5 3.2 775 2 B72074
37 106.5 3.2 775 2 C81594
38 106.5 3.2 775 2 D86549
39 106.5 3.2 832 2 H84848
40 105.5 3.2 1611 2 T38236
41 105 3.2 499 2 T45749
42 104 3.2 509 2 D87452
43 104 3.2 585 2 S72824
44 104 3.2 986 2 T33135
45 104 3.2 1917 2 C88728

ALIGNMENTS

RESULT 1

E81385

probable oxidoreductase chain Cj0415 [imported] - Campylobacter jejuni (strain NCTC 13637)
C:Species: Campylobacter jejuni.
C:Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 31-Mar-2000
C:Accession: E81385
R:Parkhill, J.; Wren, B.W.; Mungall, K.; Ketley, J.M.; Churcher, C.; Basham, D.; Chli

C.W.; Ouali, M.; Rajandream, M.A.; Rutherford, K.M.; VanVleet, A.; Whitehead, S.; B
Nature 403, 665-668, 2000
A:Title: The genome sequence of the food-borne pathogen Campylobacter jejuni reveals
A:Reference number: A81250; MUID:20150912
A:Accession: E81385

A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-573 <PAR>
A:Cross-references: GB:AL139075; GB:AL111168; NID:g6967817; PIDN:CAB74251.1; PID:g6967817
A:Experimental source: serotype O2, strain NCTC 11168
C:Genetics:
A:Gene: Cj0415

Query Match 6.1%; Score 200.5; DB 2; Length 573;
Best Local Similarity 20.6%; Pred. No. 3.5e-07;
Matches 135; Conservative 74; Mismatches 270; Indels 175; Gaps 26;

QY 36 VDVFTAGSPICATYAKLCVEAGLRVWVEIG---AADSFYAVNAEGTAVPVGYHKK 92

Db 8 VDVVTGAGTGGIIVAAELTKAGLVLSLRGHMQSTENFYTHDEWRYGINY----- 60

QY 93 NEIEFQKOIDRFVNVIKGALQOVS---VPVRNONVPTLDPGAWSPGSSAISNGKNPHQ 149

Db 61 -----GLMQDCSKDTVFRH-----DPSGLALP----- 83

QY 150 REFENLSAEVTRGVGGMSTHTWCTSPRIHPPMESLPGIGRP-----KLSNDPAEDD----- 201

Db 84 --YRKMGSFLLGNNGVGGAGVHNGWTFRPMYPDFEIQTLISKORYKNGKLDNYTLQDWGVT 141

QY 202 KENNELYSEARLIGTSTKEF-----DESIRHTVLRSLODAYKQRORI 245

Db 142 YKDMPEYDRFETKCTGCVGSEPNLAEKMGAFRSSPYQPPELNTKMLKRFESAASNNLH 201

QY 246 FRPLPLA-----CHRLKNAPEVVEHSAENLHFSIYNDKQKK 283

Db 202 TYRLPASNSKGYTPDGDODLAPCOYCAVCERFG-----CEYGAKASPLNTVTPKAMSTG 256

QY 284 LFTLTNRHRCRLAHTGGYEKKIGAAEVNLLATRNPSQLDSYIM-AKVYVLASGAIGN 342

Db 257 KYTIRYSNVTQIL-----KKDGKVTGVKFDVDTRI-----MKEYIQPADIVILVTSMENN 306

QY 343 POILYNSFGSLQVTP-----RNDSLIPNLG--RYITEQPMAPCQVLKQEFVDSVR 392

Db 307 AKLLMVSNI-GEQYDPKTKGTLGRNYCYQMNGMTTAFDFEQNTF-----MGSGALGTTTS 361

Db 235 QYRNLGIRGCPFGAYSSNSG-----GLIAARTGNLVRPNSIVTTELLIYDERAGRASGV 290
QY 313 NLLATRPSSOLDYSIYAKVYVVLASGAIGNPQILYNSGSLQVTPRNDLSLIPN----- 366
Db 291 RIL---DAETRKDEEFHADVIFLCASALNSAWIMNS-----TSRFPNGFGNAS 337
QY 367 --LGRYTEQPMAPFCQIVLRQEFVDSV---RDDPYGLPWKKEAQAHIKNTPTDALPIP 420
Db 338 DOLGRNVMDHILGAGATGCAPEFADMTYFSRRPNNGIYVPRFN--LGDAASKRSLYLRGFG 396
QY 421 FRDEPQVPTTFTEHPWHTQIHRDAFSYGAV-----GPEVDSRVIVDLRMFGATDP 472
Db 397 YQGAGRAT-----WERDRGGGRGFGAARKAALSQPGWT-----HGLSGFGEMLP 443
QY 473 EANNLLVFQNDVQDGYSMPOPTFRYRSTASNVRAKMMADMCEVASNLGYLPTSPQF 532
Db 444 YADRNVTLNRDVEDKFGPLTTLTNNVTNRDNEAMRRDQAAAAEMLAAGFQNVRAHDNG 503
QY 533 MDPLALHLAGTTRIGFDKATTVADNNSLYWDFANLYVAGNGTIRTGTGFGNPILTSMCHA 592
Db 504 FAPGLGHEGTARGHRDPTKTVLNHNVHECKNVITVTDGAAMASASCVPNSLTLYMALT 563
QY 593 IKSARSINTLKG 606
Db 564 ARAADHAVRAKRG 577

RESULT 4

AB2003
dehydrogenase chain [imported] - Anabaena sp. (strain PCC 7120)
C:Species: Anabaena sp.
A:Note: Anabaena sp. (strain PCC 7120) is a synonym of Nostoc sp. strain PCC 7120.
C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 11-Jan-2002
C:Accession: AB2003
R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S.
DNA Res. 8, 205-213, 2001
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Anabaena sp. (strain PCC 7120)
A:Reference number: AB1807; MUID:21595285; PMID:11759840
A:Accession: AB2003
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-619 <KUR>
A:Cross-references: GB:BA000019; PIDN:BA077942.1; PID:g17135396; GSPDB:GN00179
A:Experimental source: strain PCC 7120
C:Genetics:
A:Gene: air1576

Query Match 5.7%; Score 187.5; DB 2; Length 619;
Best Local Similarity 19.7%; Pred. No. 3.9e-06;
Matches 139; Conservative 80; Mismatches 242; Indels 243; Gaps 29;
QY 29 ETYNGDGVDFIAGSGPIGATYAKLCVEAGLRVVMVEIGAADSFYAVNABEGTAVPYVG 88
Db 12 DTVSKTVDAVIVGTGVAGAIKALSOQCKRILLIATV----- 51
QY 89 YHKNEIE-FQKIDIRFVNVIKALQOVSVPRNQNVPTLDPGAWSPGSSA----- 140
Db 52 -HKDLTLAGFSYVDTFYKAVD-----KNPNSP-----YPANSNVQSPTDYN 92
QY 141 ---ISNGKNPHQREFENLSAEATRGVGMSTHTWCSTPRIHP---PMESLPIGRPKLS 194
Db 93 DYIEQGPMP-----LAGSTRVLGGTTHWEAKTPMLPDEPKLSYGOGL----- 140
QY 195 NDAEDDKENNELYSEARLIGTSTK-----EFDESIRHTLVLSQDAKDRORI 245
Db 141 -DWPIDYHDLPEYRYRAEHMGVCGVDQORALGLEFPQD--YVFPMEKLPSPSYLOKVI 197
QY 246 -----FRPLPLACHRLKNAPEYVHNSAENLF-----HSTYND 279
Db 198 EKVNGTNVELYKHTLTLSFTFPQARNGVN-PKY-----DOGNLFVDPDGVTSVHPVOYGE 252

QY 280 K-----QKKLFILLTNRCTRRLALTGGYEKKIGAAEVRNLLA 316
Db 253 RCOGNANCVPCPVQAKYDARRTLKAFETGKVHLYQAVAYKVEYDQTGRTAIHYKH 312
QY 317 TRNPSSOLDYSYIMAK--VYVLASGAIGNPQILYNSGSLQVTPRNDLSLIPNLGRYITEQ 374
Db 313 YKPNSEYTTGIAKGTLFVLATNVAENARLLGSDL-----PNTSLI---GRYL--- 360
QY 375 PMAFQCQIVLRQEFVDSVRDDPYGLPWKKEAQAHIKNTPTDALPIPPRDE-----PQ 427
Db 361 -----MDHPFTLAW-----ALM-----PEVTGDMRGPL 383
QY 428 VTTTFF-----TEHPWHTQIHRDAFSYGAVGPEVDSRVIVD----- 463
Db 384 VTSIGTFRKGDPRKKOSAFAYDIHNDGWGWTATGSPKSEVEDAYDNKNKYQELRQTLIS 443
QY 464 -----LRWFGATDPEANLLVFQNDVQDGYSMPOPTFRYRSTASNV---RARKMMADM 514
Db 444 RISQOLLAFWCELLPEYGNRVITDPRHKKLG-----NYRPVINFNLDPYSRRTLAYT 497
QY 515 CEVA-----SNLGG-----YLTPSPQMDPGLAL-----HLAGTTRIGFDKATTVADNN 559
Db 498 RKVSRVIFERLGAEDYTHYDPODPAYFEFEGEYVYKGNHFSGTHIMGTPLNSVVDY 557
QY 560 SLVWDFANLYVAGNGTIRTGTGFGNPILTSMCHAISARSINTL 603
Db 558 LRSWDHKNLFLVAGSMPTIGSSNTTLTIALSPRTAHLMLQEL 601

RESULT 5

H83362
gluconate dehydrogenase PA2265 [imported] - Pseudomonas aeruginosa (strain PA01)
C:Species: Pseudomonas aeruginosa
C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C:Accession: H83362
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; Adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pathogen.
A:Reference number: A82950; MUID:20437337
A:Accession: H83362
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-591 <STO>
A:Cross-references: GB:AE004652; GB:AE004091; NID:g9948287; PIDN:AA050563.1; GSPDB:G1
A:Experimental source: strain PA01
C:Genetics:
A:Gene: PA2265

Query Match 5.5%; Score 182; DB 2; Length 591;
Best Local Similarity 20.6%; Pred. No. 9.7e-06;
Matches 138; Conservative 66; Mismatches 261; Indels 204; Gaps 23;

QY 36 VDVFIAGSGPIGATYAKLCVEAGLRVVMVEIGAADSFYAVNABEGTAVPYVGHKKNEI 95
Db 8 VDAVVVVGFGTGAAILAKELTEAGLVVALERGNRDYTPDGAYPNTL-----DEL 57
QY 96 EFQKIDIRFVNVIKALQOVSVPRNQNVPTLDPGAWSPGSSAISNGKNPHQREFENL 155
Db 58 TYNTRKLFONLSKSTV-----SIRHGINDTALPYRQL 90
QY 156 SABEATRGVGMSTHTWCSTPRIHPMESLPG-----IGRPKLSNDPAEDDKENW 205
Db 91 SAFPGLDGGVSGAGLHNSGVHFRIMPELRLSHYEERYKKFIPEGMTIQDYGVSYSELE 150
QY 206 ELYSEARLIGTSTKFEDESIRHTL-----VLSLQDAYKDRORIF 246
Db 151 PHDFEAKVEGTS-----GTAHTVKQVYKGNPFAADRSDDFPLPALRQVYS--AQLF 202
QY 247 R-----PLPLA-----CHRLKNAPEYVHNSAENLFHSIY 276

```

Db 343 DKLGITFVKA-----ALNLSPEH-----YADAKAGHAGFGDQTGAQKNTLTLFLKDAFEAG 393
QY 287 --LLTNHRCRLALTGGYEKKIGAAEVRNLLATRNSSQLDSYIMAKVYVVLASGAIGNPQ 344
Db 394 ARILVTRAQIRVLVDG-----RAAGVSATVTMGDETROIT--VRAPOVVVYAGGALETPA 446
QY 345 ILYNSGFGSLQVTPRNDLSLIPNLGRVITEQPMACFOILVRQEFVDSVRDDPYG---LPWW 401
Db 447 LLLRSIGG-----PAAGRYLRHLPAGLVAGI-----YGEDQRAWW 482
QY 402 KEVAQHIANKPTDALPIPRDE-----PQVTT--PFT--EEH-PWHTQI 442
Db 483 GP-----POSGILKOFADHENGHGFIEGVQYCPALMASGLPWTGGEAHRDLMSKF 533
QY 443 HRDAFSYGAVPEVDSRVIVD-----LRWFGATDP--EANNLLVFONDVODGYSM-----P 491
Db 534 HRWATFVSIVODRGHGVTVDDGNAVHTYALTDLDLARN---FRGVTESIRLHEAAGA 590
QY 492 QPTFRYRPSTASNVARKMADCEVAS---NLGGYLPTSPPOFMDPGLALHLAGTTRIG 548
Db 591 EEIVALAPGVAPWRGDDLEAFIGQVAPVPLGAGGQTVES-----AHQMGSAHWG 640
QY 549 FDKATTVDNNSLVDFANLYVAGNTIRTFGENPTLTSMDCHAIKASRIINTLKGTTD 608
Db 641 SDPQTSVADPDGQLHDVPGVWIGDTSFAFTPCSGVNPVSCMALASRTAEKLLAAMEGADG 700
QY 609 GKNTG 613
Db 701 GTGSG 705

```

```

RESULT 8
JC7628
glucoside 3-dehydrogenase (EC 1.1.99.13) - Halomonas sp. alpha-15
C:Species: Halomonas sp. alpha-15
C:Date: 30-Jun-2001 #sequence_revision 30-Jun-2001 #text_change 30-Jun-2001
C:Accession: JC7628; PC7123
R:Kojima, K.; Tsugawa, W.; Sode, K.
Biochem. Biophys. Res. Commun. 282, 21-27, 2001
A:Title: Cloning and expression of glucose 3-dehydrogenase from Halomonas sp. alpha-15
A:Reference number: JC7628; MUID:21164693; PMID:11263965
A:Accession: JC7628
A:Molecule type: DNA
A:Residues: 1-562 <KOJ>
A:Accession: PC7123
A:Molecule type: protein
A:Residues: 2-11 <KO2>
C:Comment: This enzyme, as an oligomeric enzyme composed of catalytic and electron trans
and disaccharides.
C:Genetics:
A:Gene: g3dh
C:Keywords: oxidoreductase

```

```

Query Match 5.2%; Score 171.5; DB 2; Length 562;
Best Local Similarity 19.9%; Pred. No. 5.7e-05;
Matches 133; Conservative 85; Mismatches 236; Indels 215; Gaps 32;

QY 37 DVFIAGSGPIGATYAKLCVEAGLRVYVMEIGAADSFYAVNAEEGTAVPYVGYHK-----91
Db 7 DAIVGSGISGGWAAKELTEKGLKVL-----ERGRNIEHVKYDHNADKEA 53
QY 92 -----KNETEFKIDREFNVNKGALQVSVVVRNVPTLDP--CAWSAPGSSAISNG 144
Db 54 WDPHPRNE-PTQMIKY-----PVLKRDYPLNEATLGMWA-----DEQ 91
QY 145 KNPH--QREFENLSAEAVTRG--VGMSTHWTCTSTPRIHPHPSLPGIGRPKLSND---P 197
Db 92 ANPYVEEKFRDWF-----RGYHVGGSLMGRQSYRLSP--MDFEANQREGIAIDWPPIR 143
QY 198 AEDDKENNELYSAEERLIG-TSTKEFDESIRHTLVLSRLQDAYKQRIFRPLPLACHRL 256

```

```

Db 144 YEDLAPW---YDYVERFAGIAGTQE-----GLDILPDG-----EFLPPIPLNC--- 183
QY 257 KNAPEVVEHSAENLPHSIYNDKQKLFLLTNHRCRLALTGGYEKKIGAAEVRNLLA 316
Db 184 -----VEDNAARI-----KEAF-----GGORHLHSRVANITQ 212
QY 317 TRNPSQDSYIMAKVYVVLASGAIGNPOILYNSGFSGLQVTPRNDLSLIPNLGRVITEQPM 376
Db 213 PRPEQNRVNCQYRNKCW-----LGCP---YGAVFSTQSATLPAAVATGNL---TLRPF 259
QY 377 AFCQIVL---RO-----EFVDS-----VRDDPYGLPW-----NKEA 404
Db 260 SIYSQVLYDKDRQARGVEVIDAETHVEHYETADVIFLNASTFTNTWILNASTADVWEGG 319
QY 405 VAQHIANKPTDALPIPR-----DPEQVTTPTTEHPWHTQIHR-----DAF 447
Db 320 LGSSSELGHNVMDHFRGASGEVGYLDKYFGRPAGFYIPRFRNVGDEQBSYVRGP 379
QY 448 SY-GAVGPEVDSRVIVDLR-----W-----FGATDPEANLLVFQNDVOD 486
Db 380 GYGAAASREGWDREIAELNIGADLKQALTOPGGWTIGMTGFGEMLPDHDNRISLDHSVRD 439
QY 487 GYSMPOTFRYRPSTASNVARKMADMBCE-----VASNLGGYLPTSPPOFMDPGL 537
Db 440 KMGLPVLSIDVELKQNERDMRDVODAVDLLEAAGVKNVKGVDGYA-----PGM 490
QY 538 ALHLACTTRIGDPKATTVADNNSLVDFANLYVAGNTIRTFGENPTLTSMDCHAIKASR 597
Db 491 GIHEMGTARMGRDPKTSVLSNHNQVNDAPNVFVDGACMTSSSCVNPSTLTMALTARAYD 550
QY 598 SIINTLKG 606
Db 551 YAVEELKRG 559

```

```

RESULT 9
T10651
hypothetical protein T5F17.20 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 15-Sep-2000
C:Accession: T10651
R:Bevan, M.; Lennard, N.; Quail, M.; Harris, B.; Rajandream, M.A.; Barrell, B.G.; Ba
submitted to the Protein Sequence Database, June 1999
A:Reference number: Z16533
A:Accession: T10651
A:Molecule type: DNA
A:Residues: 1-748 <BEV>
A:Cross-references: EMBL:AL049917; GSPDB:GN00062; ATSP:T5F17.20
A:Experimental source: cultivar Columbia; BAC clone T5F17
C:Genetics:
A:Gene: ATSP:T5F17.20
A:Map position: 4
A:Introns: 70/3
C:Superfamily: Arabidopsis thaliana hypothetical protein T5K18.160

```

```

Query Match 4.8%; Score 158.5; DB 2; Length 748;
Best Local Similarity 19.4%; Pred. No. 0.00089;
Matches 126; Conservative 90; Mismatches 215; Indels 217; Gaps 28;

QY 37 DVFIAGSGPIGATYAKLCVEAGLRVYVMEIGAADSFYAVNAEEGTAVPYVGYHKNEIE 96
Db 239 DAVVVGSGGGVAAANLAKAGLVLEKG---NYTADYSGLVPSMLELYEKG---292
QY 97 FKQIDIRFNVNKGALQVSVVVRNVPTLDPGANSAPGSSAISNGKPHOREFNLS 156
Db 293 -----GLTTV-----DGL-----FMLLA 306
QY 157 AEAVTRGVGSMTHWTCTSTPRIHPHPSLPGIGRPKLSNDPAEDDKENNELYSAEERLIG 216
Db 307 GSAVG---GGTAVNWSAS---INTPDHVL-----QEW-----SEGSKIKF 340
QY 217 TSTKEF---DE-SIRHTLVLSRLQDAYKQRIFRPLPLACHRL-----KNAPE-- 261

```


Db 109 VEH-----EDGVSPANVSAEIEPPYAEAEERLEGV-ROGAGDDTPEPPRSPAPYMH 158

QY 209 SEA---ERLIGTSTKEFDESIRHTLVLSLSDAYKDRORIFRPLPLA-----CHRLKN 258

Db 159 APIPEHPVIGRVAKGFER-----LGLRPFH-----MPSAIDYVGGGLCHRCGT 201

QY 259 APEYVEVWSEANLPHSYINDDKOKKLF-----TLLTNHRCRTALHTGGYEKKIGAA 309

Db 202 CDAFVCRFDAGK-----DAETRLRLPALRHPNVSLLTGARVRL- IADSGDGKHIVAV 252

QY 310 EVRNLLATRNPSOLDISYIMAKVVVLASGAIGNPOIILYNSGFSGLQVTPRNDLSLPN--- 366

Db 253 EIER-----AGEITT-IEAPLEVLSAGAINSALLILRSA-----DEKKPNGLA 294

QY 367-----LGRYITEQPM-----FCQIVLRQEEVSDVRDDP----- 395

Db 295 NSSGVVGRYLMNHLISGLMGLLPFTINDTRFPKTMSLNDFDGTGDEAARGNVQMLGNI 354

QY 396-----YGLPMWKCAVAQOHTAKNPTDALPTPFRRDPEQVTPPTTEHPHHTQIHRDAFSY 449

Db 355 QGPMIRAAYPMWPRPLANLLARHSVDFLVM-----SED----- 387

QY 450 GAVGPEVDSRVIVDLRWFGATDPEANLLVFQNDVQDVSMPQPTFRYRPSTASN---V 505

Db 388---TPKKVDSRV-----KPGKNGAE-----LIYRPGDEAHORFV 419

QY 506 RARKWADMCVEASNLGGLYTPSPQPMDPGLALHLAGTTRIGFDKATTVAADNSLVWDF 565

Db 420 RHMSLLRKKNGFPVVLGHSGFIEAPS-----HCGGIVRMGDDPKAAALNALCQTYDH 471

QY 566 ANLYVAGNGIIRTGFGNGENPTLTSCHAIK 594

Db 472 PNLVYVDAGFPSPSAALNPALTVAQAALR 500

RESULT 13

AD3084

dehydrogenase Atu4296 [Imported] - Agrobacterium tumefaciens (strain C58, I

C:Species: Agrobacterium tumefaciens

C:Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 11-Jan-2002

C:Accession: AD3084

R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Ch

erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutayavin, T.; Levy, R.; L

; Karp, P.; Romero, P.; Zhang, S.

Science 294, 2317-2323, 2001

A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.;

ster, E.W.

A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaci

A:Reference number: AB2577; PMID:11743193

A:Accession: AD3084

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-527 <KUR>

A:Cross-references: GB:AE008689; PIDN:RAL45090.1; PID:gl17742757; GSPDB:GN00

A:Experimental source: strain C58 (Dupont)

C:Genetics:

A:Gene: Atu4296

A:Map position: linear chromosome

Query Match 4.23; Score 138.5; DB 2; Length 527;
Best Local Similarity 19.18; Pred. No. 0.017;
Matches 120; Conservative 74; Mismatches 222; Indels 213; Gaps 25;

Matches	120;	Conservative	74;	Mismatches	222;	Indels	213;	Gaps	
Qy	37	DVFITAGSPIGATYAKLCVEAGLRVVMVEIGAADSFYAVNAEGTAVPVVPYGHKKNEIE	96						
		: :	:	: : :	:	:	:	:	:
Db	14	DIVITSGVGGSVALKLAATGAKILLIERGEKLPMENPAD--AEAFFVQNRYTTDL-	70						
		: :	:	: : :	:	:	:	:	:
Qy	97	FQKDIDRFVNVIKGALQQVSVPPVRNNONFTLPDQAWSAPP-----SSAISNGKNPHORE	151						
		: :	:	: : :	:	:	:	:	:
Db	71	-----YRDTGTPTPGQYYVGGHTKYCTGTAMFRFDRDRPRE	108						
		: :	:	: : :	:	:	:	:	:
Qy	152	FENLSAEAVTRGVGGMSTHWTCSTPRIHP----PMSLPGIGRPKLNSDPADDDKEWNELY	208						
		: :	:	: : :	:	:	:	:	:

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 26, 2002, 11:03:42 ; Search time 14.68 Seconds
(without alignments)
1028.271 Million cell updates/sec

Title: US-09-856-327-2
Perfect score: 3284
Sequence: 1 MSLSTEQLRDPYPRSMQING.....IINTLKGTGDKNTGEHRNL 618

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents_AA:*

1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep.*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep.*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep.*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep.*
5: /cgn2_6/ptodata/2/iaa/PTUS_COMB.pep.*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query	Score	Match	Length	DB ID	Description
1	1232.5	37.5	623	1	US-08-734-925-2	Sequence 2, Appli
2	1210	36.8	622	4	US-09-305-381-2	Sequence 2, Appli
3	1091	33.2	566	4	US-09-023-731-1	Sequence 1, Appli
4	192	5.8	51	4	US-09-023-731-12	Sequence 12, Appli
5	191.5	5.8	615	4	US-09-297-937C-9	Sequence 9, Appli
6	191.5	5.8	1276	4	US-09-297-937C-13	Sequence 13, Appli
7	130.5	4.0	58	4	US-09-023-731-11	Sequence 11, Appli
8	118.5	3.6	785	3	US-09-265-108-2	Sequence 2, Appli
9	118.5	3.6	785	4	US-09-479-264-2	Sequence 2, Appli
10	115	3.5	50	4	US-09-023-731-14	Sequence 14, Appli
11	108	3.3	35	4	US-09-023-731-15	Sequence 15, Appli
12	104.5	3.2	1876	2	US-08-609-049A-12	Sequence 12, Appli
13	104.5	3.2	1876	4	US-09-170-996-12	Sequence 12, Appli
14	102.5	3.1	61	4	US-09-023-731-6	Sequence 6, Appli
15	100.5	3.1	806	3	US-08-549-515-5	Sequence 5, Appli
16	100.5	3.1	806	3	US-08-549-515-11	Sequence 11, Appli
17	100	3.0	664	1	US-08-485-284A-5	Sequence 5, Appli
18	98.5	3.0	543	3	US-09-199-229-2	Sequence 2, Appli
19	98.5	3.0	543	4	US-09-443-087-2	Sequence 2, Appli
20	98.5	3.0	543	4	US-09-687-298-2	Sequence 2, Appli
21	98	3.0	1257	1	US-08-049-783-2	Sequence 2, Appli
22	98	3.0	1257	1	US-08-158-232-6	Sequence 6, Appli
23	98	3.0	1257	1	US-08-304-626-6	Sequence 6, Appli
24	98	3.0	1257	1	US-08-316-301A-6	Sequence 6, Appli
25	98	3.0	1257	2	US-08-611-928-6	Sequence 6, Appli
26	98	3.0	1257	3	US-09-173-891-6	Sequence 6, Appli
27	98	3.0	1257	4	US-09-076-137-6	Sequence 6, Appli

Sequence 6, Appli
Sequence 5, Appli
Sequence 5, Appli
Sequence 28, Appli
Sequence 28, Appli
Sequence 2, Appli
Sequence 22, Appli
Sequence 22, Appli
Sequence 1, Appli
Sequence 1, Appli
Patent No. 5244792
Sequence 32, Appli
Sequence 30, Appli
Sequence 2, Appli
Sequence 14, Appli
Sequence 2, Appli
Sequence 6, Appli
Sequence 6, Appli

28 98 3.0 1257 5 PCT-US92-03624-6
29 97 3.0 3170 2 US-07-642-734C-5
30 97 3.0 3170 3 US-08-439-009A-5
31 96.5 2.9 1876 2 US-08-609-049A-28
32 96.5 2.9 1876 4 US-09-170-996-28
33 95.5 2.9 382 1 US-08-186-833-2
34 93.5 2.8 903 3 US-08-804-439A-22
35 93.5 2.8 903 3 US-08-720-229-22
36 93.5 2.8 904 4 US-08-632-537-1
37 93.5 2.8 904 5 PCT-US96-05316-1
38 93.5 2.8 904 6 5244792-4
39 93 2.8 604 2 US-08-746-283-32
40 93 2.8 604 2 US-08-746-257A-30
41 93 2.8 605 1 US-08-333-802-2
42 93 2.8 4545 2 US-08-804-227C-14
43 92.5 2.8 4472 2 US-08-804-227C-2
44 92 2.8 844 1 US-07-731-157A-6
45 92 2.8 844 2 US-08-541-780-6

ALIGNMENTS

RESULT 1
US-08-734-925-2
; Sequence 2, Application US/08734925
; Patent No. 5712139
; GENERAL INFORMATION:
; APPLICANT: NISHIMURA, IKUKO
; APPLICANT: OKADA, KIMIHARU
; APPLICANT: MINAMIHARA, TOMOYUKI
; APPLICANT: KAWAI, GENSHIRO
; APPLICANT: KAYAMA, YASUJI
; APPLICANT: SUZUKI, MASARU
; TITLE OF INVENTION: NOVEL PYRANOSE OXIDASE, PYRANOSE OXIDASE
; TITLE OF INVENTION: GENE, NOVEL RECOMBINANT DNA AND PROCESS FOR PRODUCING
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,
; ADDRESSEE: P.C.
; STREET: 1755 S. JEFFERSON DAVIS HIGHWAY, SUITE 400
; CITY: ARLINGTON
; STATE: VA
; COUNTRY: USA
; ZIP: 22202

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/734,925
FILING DATE: 22-OCT-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/568,428
FILING DATE: 06-DEC-1995
ATTORNEY/AGENT INFORMATION:
NAME: OBLON, NORMAN F.
REGISTRATION NUMBER: 24,618
REFERENCE/DOCKET NUMBER: 7126-002-0
TELECOMMUNICATION INFORMATION:
TELEPHONE: 703-413-3000
TELEFAX: 703-413-2220
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 623 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-734-925-2

MEDIUM TYPE: FLOPPY DISK
COMPUTER: IBM PC COMPATIBLE
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: MICROSOFT WORD 97
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/023,731
FILING DATE: 13-FEB-1998
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 29275/1997
FILING DATE: 13-FEB-1997
ATTORNEY/AGENT INFORMATION:
NAME: KENNETH H. SONNENFELD
REGISTRATION NUMBER: 33,285
REFERENCE/DOCKET NUMBER: 3479-4000
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 758-4800
TELEFAX: (212) 751-6849
TELEX: 421792
INFORMATION FOR SEQ ID NO: 11:
SEQUENCE CHARACTERISTICS:
LENGTH: 58
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-09-023-731-11

Query Match 4.0%; Score 130.5; DB 4; Length 58;
Best Local Similarity 44.4%; Pred. No. 3.1e-06;
Matches 28; Conservative 11; Mismatches 13; Indels 11; Gaps 2;

QY 333 YVLASGAINPQILYNSGFGSLQVTPRNDLIPNLGRYITEQPMATCQIVLRQEFVDSVR 392
DB 2 FVACGAVCTPQILWS-----NIRPY-----ALGRYLSQSMTCQIVLRGIYDATA 50
QY 393 DDP 395
DB 51 TDP 53

RESULT 8
US-09-265-108-2
; Sequence 2, Application US/09265108
; Patent No. 6033891
; GENERAL INFORMATION:
; APPLICANT: Golightly, Elizabeth
; TITLE OF INVENTION: Nucleic Acids Encoding Polypeptides
; FILE REFERENCE: 5850.000-US
; CURRENT APPLICATION NUMBER: US/09/265,108
; EARLIER FILING DATE: 1999-03-09
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 785
; TYPE: PRT
; ORGANISM: Humicola
US-09-265-108-2

Query Match 3.6%; Score 118.5; DB 3; Length 785;
Best Local Similarity 19.5%; Pred. No. 0.0055;
Matches 126; Conservative 93; Mismatches 257; Indels 171; Gaps 30;

QY 26 AIHETGNDGVDFIAGSGPIGATYAKLCVEAGLRVYVMEIGAASFYAVNAEEGTAVP- 84
DB 242 APEETY-----DYIVVGAGAGGIPVADKLSAGHKVLLIEKGPP-----STGRWQGTMKPE 292
QY 85 YVPGYHKKNEIEFQKIDIRFVNVIKALQOVSVPVNRQNVPTLDPGMSAPPGSSAISNG 144
DB 293 WLEG-----TDLTRF--DVPGLCNIWVDSAGIACITDQMGAGCVLGGGTAVNAG 340

QY 145 K--NPHOREFENLSAEAVTRGVGGMSTHTWCSTPRHHPMESLPGIGRKLNSDPAED-- 200
DB 341 LWWKPIDLDWDENFPE-----GWSQDLAAATERV---FERIEGTWHPSDMGKLYRDEG 391
QY 201 -----DKWNELYSEARELIGTSTKEFDESIRHTLVLSRQDAYKQRIRPLPL 251
DB 392 YKVLSSGLAESGWKEV-----VANEVPEKNKRTFAHTHEMFAGGE-----RNGPL 436
QY 252 ACHRLKNAPEYVWEHSAENLFHSIYNDKOKKLTLLTNHRCRLATLTGGYEKKIGAAEV 311
DB 437 ATY-----LVSADAREN-----FSLWTNTAVRRAVETGG---KVTGVEL 472
QY 312 RNLATRNPFSSQDLSYIMAKVYVVLASGAIGNPOILYNSGFGSLQVTPRNDLIPNLGRYI 371
DB 473 ECL--TDGGYSGIVKLNKNEGVIFFSAGAFSAKLLFRSGIG----- 511
QY 372 TEOPMAFCQIVLR-----QEFVDSVRDDPYGLPMWKEAVAQHIKKN-PTDAL-----PIPF 422
DB 512 ---PEQQLRVVASSKDGEDFIDE-KD-----WIKLPVGYNLIDHLNTDLTLTHPDVVVF 561
QY 423 DPEPQVTPTEEPHWHQIHRDAFSYGA--VGP-----EYDSRVIVDL 464
DB 562 DFYEAWTTPIEADKQILYLEORSIGILAAQAPNIGPMWEEQVTPSDGITRQFQWTA RBGDS 621
QY 465 RMEGATDPEANLLVFQNDVQDCYS-----MPQPTFRYRPSTASNVRAKMMADM 514
DB 622 RFTNSSHAMTSLQYLGKGVVSRGRATITQGLVTVVAEHPYLHNAGDKAEVIOGKNLIES 681
QY 515 CEVASNLGGVLPSP-----QFMDPGL-----ALHLAGTTRIGFDK-----ATTVADN 558
DB 682 LNVIPNITWLP--PPGSTVEEYVDSLLVSASARRSNHMMGTAKLGTDDGRYGTGSVVDL 739
QY 559 NSLVWDFANLYVAGNCTIRGTGCE-NPTLTSMCHAKSARSINTLK 604
DB 740 DTKVCTDNLVV--DASIPFGMSTGNPSAMIVIAAEQAERIILKLR 785

Query Match 3.6%; Score 118.5; DB 4; Length 785;
Best Local Similarity 19.5%; Pred. No. 0.0055;
Matches 126; Conservative 93; Mismatches 257; Indels 171; Gaps 30;
QY 26 AIHETGNDGVDFIAGSGPIGATYAKLCVEAGLRVYVMEIGAASFYAVNAEEGTAVP- 84
DB 242 APEETY-----DYIVVGAGAGGIPVADKLSAGHKVLLIEKGPP-----STGRWQGTMKPE 292
QY 85 YVPGYHKKNEIEFQKIDIRFVNVIKALQOVSVPVNRQNVPTLDPGMSAPPGSSAISNG 144
DB 293 WLEG-----TDLTRF--DVPGLCNIWVDSAGIACITDQMGAGCVLGGGTAVNAG 340
QY 145 K--NPHOREFENLSAEAVTRGVGGMSTHTWCSTPRHHPMESLPGIGRKLNSDPAED-- 200

RESULT 13
 US-09-170-996-12
 ; Sequence 12, Application US/09170996
 ; Patent No. 6291220
 ; GENERAL INFORMATION:
 ; APPLICANT: Williams, Lewis T.
 ; APPLICANT: Molz, Lisa
 ; APPLICANT: Chen, Yan-Wen
 ; TITLE OF INVENTION: No. 6291220el PI 3-Kinase Polypeptides
 ; NUMBER OF SEQUENCES: 32
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Townsend and Townsend and Crew LLP
 ; STREET: Two Embarcadero Center, 8th Floor
 ; CITY: San Francisco
 ; STATE: California
 ; COUNTRY: USA
 ; ZIP: 94111-3834
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/09/170,996
 ; FILING DATE:
 ; CLASSIFICATION: 435
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 08/609,049
 ; FILING DATE: 29-FEB-1996
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Dow, Karen B.
 ; REGISTRATION NUMBER: 29,684
 ; REFERENCE/DOCKET NUMBER: 2307K-0637000US
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 415-326-2400
 ; TELEFAX: 415-326-2422
 ; INFORMATION FOR SEQ ID NO: 12:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 1876 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 US-09-170-996-12

```

Query Match      3.28; Score 104.5; DB 4; Length 1876;
Best Local Similarity 22.08; Pred. No. 0.72;
Matches 85; Conservative 53; Mismatches 170; Indels 79; Gaps 21;

Qy      83 VPYYPGGHKKNEIEFQKIDRFVNVWIGALQQVSVPRVQNVPVLPDGA----NSAP--P 136
      |||  ::  ::  ::  ::  |||  ::  |||  ::  |||  ::  |||  ::  |||  ::
Db      184 VPYQAAQQQORPLN--SEELQRLYSN--PAQMAVVPVPPQPNAYMYPGAVVTPYTAIYP 239

Qy      137 GSSAISNGKNPHREFENLSAEATRVGGVGHSTHWTCTSTPRIHP--PMESLPGIG--RPKL 193
      |||  ::  ::  ::  ::  |||  ::  |||  ::  |||  ::  |||  ::  |||  ::
Db      240 GSAAFMPPQYPAQ-----GYGFGGAYTHMDLRRRQSQPAPQOTAPTTHSHHSQP 288

Qy      194 SNDPAEDDKENNELYSEAR----LIGSTKEFDESIRHTLVLSLQDA----YKDRQRI 245
      |||  ::  ::  ::  ::  |||  ::  |||  ::  |||  ::  |||  ::  |||  ::
Db      289 SNHSTSPAEANGVAFFARQVPSTVGVSSSHSTGNGNGSHSVPRRGNDLIDLNHEDYSRV 348

Qy      246 ----FRPLPLACHRLKNAPYVSWHAENLFHSIYNDKDKKFLTLLTNHRCRLALTG 300
      |||  |||  |||  ::  ::  ::  ::  |||  ::  |||  ::  |||  ::  |||  ::
Db      349 SVLEAFDPLNDNTGNTADSTSYTAYEYDFDELYSGD-----AATQSDPMYEAVN 401

Qy      301 GYEKKIGAAAEVR--NLLATR-----NPSSQLDSYTYMA-----KYVVLASGAIGNP----- 343
      ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::
Db      402 RWDYTVATVSPNVLGWRQDFLQSPSTSSSQYGVAPPESLKLAENSGTISPPPLPP 461

Qy      344 --QILYNSGFGSLQVTF--PRNDSLIPNLGRITYTEOPMAFCQIVLRQ-----EFVDSVR 392
      ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::

```


GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 26, 2002, 11:03:12 ; Search time 34.15 seconds
(without alignments)
2010.062 Million cell updates/sec

Title: US-09-856-327-2

Perfect score: 3284

Sequence: 1 MSLSTEQLRDRPRSMQING.....IINTLKGGTGDGKNTGEHRL 618

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- 1: A_Geneseq_032802.*
- 2: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1980.DAT.*
- 3: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1981.DAT.*
- 4: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1982.DAT.*
- 5: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1983.DAT.*
- 6: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1984.DAT.*
- 7: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1985.DAT.*
- 8: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1986.DAT.*
- 9: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1987.DAT.*
- 10: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1988.DAT.*
- 11: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1989.DAT.*
- 12: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1990.DAT.*
- 13: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1991.DAT.*
- 14: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1992.DAT.*
- 15: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1993.DAT.*
- 16: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1994.DAT.*
- 17: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1995.DAT.*
- 18: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1996.DAT.*
- 19: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1997.DAT.*
- 20: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT.*
- 21: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT.*
- 22: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA2000.DAT.*
- 23: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Score	Length	ID	Description
1	3284	100.0	618	22	Lyophyllum shimeji
2	1232.5	37.5	623	20	Coriolus versicolor
3	1232	37.5	633	21	Pleurotus cornucop
4	1218.5	37.1	623	17	Pleurotus ostreatus
5	1210	36.8	622	22	Pyrenopeziza tritici
6	1091	33.2	566	20	Trametes hirsuta p
7	1044	31.8	564	21	An antitumour prot
8	1044	31.8	564	21	T. matsutake pyran
9	200	6.1	544	20	Trichoderma derm
10	192	5.8	51	20	G. oxydans D-sorbi
11	191.5	5.8	615	20	Peptide derived fr
					Gluconate dehydrog

12	177	5.4	67	21	AA10459
13	177	5.4	67	21	AA10459
14	168	5.1	34	22	AA10459
15	162	4.9	30	22	AA10459
16	159.5	4.9	657	21	AA10459
17	159.5	4.9	675	21	AA10459
18	159.5	4.9	748	21	AA10459
19	142	4.3	29	22	AA10459
20	130.5	4.0	58	20	AA10459
21	131	3.7	24	22	AA10459
22	121	3.7	704	20	AA10459
23	120.5	3.7	546	20	AA10459
24	119	3.6	500	22	AA10459
25	119	3.6	545	21	AA10459
26	119	3.6	727	21	AA10459
27	119	3.6	746	21	AA10459
28	119	3.6	2639	22	AA10459
29	118.5	3.6	785	21	AA10459
30	118.5	3.6	785	22	AA10459
31	116.5	3.5	698	20	AA10459
32	115	3.5	50	20	AA10459
33	110	3.3	551	21	AA10459
34	108	3.3	35	20	AA10459
35	107.5	3.3	964	22	AA10459
36	106.5	3.2	778	20	AA10459
37	106	3.2	572	22	AA10459
38	106	3.2	1356	22	AA10459
39	105	3.2	609	22	AA10459
40	105	3.2	833	22	AA10459
41	104.5	3.2	1876	18	AA10459
42	104.5	3.2	1876	22	AA10459
43	104.5	3.2	1876	22	AA10459
44	103	3.1	437	22	AA10459
45	103	3.1	1686	19	AA10459

ALIGNMENTS

RESULT 1

ID AAB97035 standard; Protein; 618 AA.

AC AAB97035;

XX 20-JUL-2001 (first entry)

DE Lyophyllum shimeji antibacterial protein.

DE Lyophyllum shimeji antibacterial protein.

OS Lyophyllum shimeji.

XX WO200121657-A1.

XX 29-MAR-2001.

XX 20-SEP-2000; 2000WO-JP06404.

XX 21-SEP-1999; 99JP-0267238.

XX (NISR) JAPAN TOBACCO INC.

XX (NOR) SOC TECHNO-INNOVATION AGRIC FORESTY & FI.

XX Takakura Y, Kuwata S, Inoue Y;

XX WPI: 2001-281598/29.

XX N-PSDB; AAF99980.

XX Antibacterial protein and encoded gene isolated from Lyophyllum

XX shimeji, with activity against plant pathogenic bacteria, applicable in

XX agriculture e.g. rice cultivation at low concentration, produced at low

T. matsutake pyran
N-terminal fragmen
Lyophyllum shimeji
Lyophyllum shimeji
Arabidopsis thalia
Arabidopsis thalia
Arabidopsis thalia
Lyophyllum shimeji
Peptide derived fr
Lyophyllum shimeji
Candida tropicalis
Choline oxidase (C
A. tumefaciens 1,5
Arabidopsis thalia
Arabidopsis thalia
Arabidopsis thalia
Novel human diagno
Humicola insolens
H. insolens DSM 18
Candida cloacae fa
Peptide derived fr
Streptomyces globi
Peptide derived fr
Amino acid sequenc
Chlamydia pneumoni
S. epidermidis ope
Drosophila melanog
Propionibacterium
Drosophila melanog
Phosphatidyl inosi
Drosophila melanog
Drosophila melanog
Novel human diagno
Human class II PI3

QY 334 VLASGAIGNPOILYNSGSGLOVTPRND-----SLIPNLGRYITEOPMAFCQIVLROEF 387
 Db 317 lltagavhnaqliaasfgql---grdpakplsllylpylghthetqlvfcqvmstel 373
 QY 388 VDSVRDDP--YGLP-----WKEAVAQHIANKPTDALPIPRDPE 425
 Db 374 insvtadmtivgkpgphdysvtytppgnnnkhpdwnekvkkmhmdhqedplpipfedpe 433
 QY 426 POVTTPTTEHPWHTOIHRAFSYGAYPEVDSRVIVDLRWFGATDPEANNLLVFQNDVQ 485
 Db 434 pqvctllqatpwhqlhrdaifsgavqgsldsrllvdwrfgrtepkkeenklwfsdkt 493
 QY 486 DGYSMPOPTFRYRSTASNVRAKRMADMCEVASNLGGYLPTSPQFMDPGLALHLAGTT 545
 Db 494 daynlrptdfdr--fpggreaedmmdmcmvsakigflpgsyppqfmgglvlhlqgth 551
 QY 546 RIGFDKAT--TVADNNSLVWDPANLYVAGNTIRTFGENPTLTSMCHAISARSIIINTL 603
 Db 552 rmgfdekadccvtdsrvgfknflggcgnlptayaantltamslaikscelyi---k 608
 QY 604 KGGTDGKNTGEHRN 617
 Db 609 knfepsnpvkhhn 622
 RESULT 6
 AAW87531
 ID AAW87531 standard; Protein; 566 AA.
 AC AAW87531;
 XX
 DT 26-FEB-1999 (first entry)
 DE An antitumour protein.
 XX
 KW Antitumour protein; Tricholoma matsutake.
 OS Tricholoma matsutake.
 PN JP10313876-A.
 PD 02-DEC-1998.
 PF 13-FEB-1998; 98JP-0031452.
 PR 13-FEB-1997; 97JP-0029275.
 XX (MOMO-) MOMOYA KK.
 PA (NORO) NORINSUISANSHO SHOKUJIN SOGO.
 DR WPI; 1999-074153/07.
 XX N-PSDB; AAW83626.
 XX
 PT An anti-tumour protein - prepared by culture of host cell
 transformed by vector containing base coding sequence
 XX
 PS Claim 1; Page 7-8; 15pp; Japanese.
 XX
 CC The present sequence encodes an antitumour protein, and is isolated
 from Tricholoma matsutake.
 XX
 SQ Sequence 566 AA;

Query Match 33.2%; Score 1091; DB 20; Length 566;
 Best Local Similarity 42.8%; Pred. No. 2.9e-93;
 Matches 248; Conservative 84; Mismatches 184; Indels 66; Gaps 15;
 QY 37 DVFIAGSGPIGATYAKLCVE--AGLRVMVMEIGAADSFYAVNAEGTAVYVPYGHKNE 94
 Db 33 dvfiagsgpiisctyarihltdntsttkvmaelgsqdn-----pvi-gahhks 79
 QY 95 IEFQKDIDRFVNVIKGALQOVSVFVRNQNVPTLDPGAWSAPPGSSA-----ISNGKNPHQR 150

Db 80 ikfkdkdkfnniingalqpispsdtyqclavaaw-appidpaegqlvimgnnpde 138
 QY 151 EFENLSAEAVTRGVGGMSTHTCTSTPRIHPHMEISLPGIGRPKLSNDPAEDDEWELYSE 210
 Db 139 aglnlpgsavrtrvggmthwtcacptphd-----eearvnpv-dkqefdaller 187
 QY 211 AERLIGSTKEPDESIRHTLVLSLDAYKQRORIFRPLPLACHRLKNAPEYVEVHSAEN 270
 Db 188 aktllnvhsdqydsdqirqlvketlqtl-dasrgvttlplqvertrdnplvywtgda- 245
 QY 271 LFHSIYNDKOKKLTLLTNHRCRLALTGGYEKKIGAEVRLNLLATRNPSQLDSYMA 330
 Db 246 ---tvlgdvkpsrfalvtetrvtklivsetnptqvvaallrnl-----ntsn--delvva 296
 QY 331 KVVYLASGAIGNPOILYNSGSGLOVTPRNDSLIPNLGRYITEOPMAFCQIVLROEFVDS 390
 Db 297 ksfviagavctpqllwns-----nirpy-----algrylseqmtfcqlvlgivda 345
 QY 391 VRDDPYGLPWKEAVAQHIANKPTDALPIPRDPEPOVTTTEHPWHTOIHRAFSYG 450
 Db 346 latdpr----faakveahkkhpdvlpipthepepvmipysdvpwhvgvhrdafsyg 401
 QY 451 AVGEVDSRVIVDLRWFGATDPEANNLLVF-----QNDVODGYSMPOPTFRYRST 501
 Db 402 dvpgkadprvvdlrffgksdiveenrvtfgnpkldreagvtdtygmptfivkrtn 461
 QY 502 ASNVRAKRMADMCEVASNLGGYLPTSPQFMDPGLALHLAGTTTRIGFDKATTVADNNSL 561
 Db 462 adgdrdqimndmtnvanmlygylpgsyppqfmgglvlhltgttrigtddqtsvadptsk 521
 QY 562 VWDFANLYVAGNTIRTFGENPTLTSMCHAISARSIIINTL 603
 Db 522 vhnfnllwvgngncipdatacnptrtsvayalkgaeavvnyl 563
 RESULT 7
 AAB10457
 ID AAB10457 standard; Protein; 564 AA.
 AC AAB10457;
 XX
 DT 11-DEC-2000 (first entry)
 DE T. matsutake pyranose oxidase protein.
 XX
 KW Pyranose oxidase; matsutake mushroom; antibacterial; sugar metabolism;
 diabetes mellitus.
 XX
 OS Tricholoma matsutake.
 PN JP2000175698-A.
 PD 27-JUN-2000.
 PF 16-DEC-1998; 98JP-0357423.
 PR 16-DEC-1998; 98JP-0357423.
 XX (NISB) JAPAN TOBACCO INC.
 XX WPI; 2000-478485/42.
 DR N-PSDB; AAB71487.
 XX
 PT A reagent containing pyranose oxidase for the determination of pyranose
 PS Claim 5; Page 17-18; 23pp; Japanese.
 CC This invention describes a novel reagent containing pyranose oxidase for
 the determination of pyranose which can be prepared from a fraction
 precipitated from an aqueous extract of matsutake mushroom by ammonium
 sulfate precipitation and has an antibacterial activity against at least

Db 169 -----eervnnpv-dkqefdalleraktlllnvhsdqydsirgqvktlqqtll-dasrgv 222
 QY 247 RPLPLACHRLKNAPEVVEVHSAENLFHSIYNDDOKKLETLTLNHRCTRLALTGGEYKXI 306
 Db 223 tllplgvertdnplyvtwagad-----twlgdvkpsprvltvtrvtkfivsetnptqv 278
 QY 307 GAAEVRNLLATRNPSQSOLDYSIMAKYVIVLASGAIGNPQILYNSGSGLOVTPRNDSLIPN 366
 Db 279 vaallnrl-----ntsn--delvvaqsvlacgavctpqllwns-----nirph-----a 321
 QY 367 LGRYITEQPMACQIVLROEFVDSVRDDPYGLPWKEAVQAQIAKNAKPTDALPIPRDPEP 426
 Db 322 lgrlyseqsmftcqlvklrksivdsiatdpr-----faakveahkkhdpddvlpipfhep 377
 QY 427 QVTTPTTEHPWHTQIHRDAFSYGAVGPEVDSRVIVDLRWFAGATDPEANNLWF----- 480
 Db 378 qvmplytsdipwlvqwhr--yatgvgpkradprvvvdlrffgksdlveenrvctfcpnpxl 435
 QY 481 ---QNDVQDGYSMPOPTFRYRPTASNVRAKMMADMCEVASNLGGLYLPSPQPMDPGL 537
 Db 436 rdweagvtdtygmppqptfhvkrtnadgdrqmmndmtanvanilgylpgsyppqfmapgl 495
 QY 538 ALHLACTTRIGFDKATVADNNSLVWDFANLYVAGNGTIRTGEGENPTLTSMCHAISAR 597
 Db 496 aqhitgttrigtddqtsvadptskvhnfdnlwvgngcipdatacnptrtsvayalkgae 555
 QY 598 SIINTL 603
 Db 556 avvsyl 561

RESULT 9
 ID AAY14049
 XX AAY14049 standard; Protein; 544 AA.
 AC AAY14049;
 XX
 DT 15-JUL-1999 (first entry)
 DE G. oxydans D-sorbitol dehydrogenase.
 XX
 KW D-sorbitol dehydrogenase; L-sorbose; 2-keto-L-gulononic acid; precursor;
 KW L-ascorbic acid production.
 XX Gluconobacter oxydans.
 XX WO9920763-AI.
 XX 29-APR-1999.
 XX 13-OCT-1998; 98WO-JP04612.
 XX 17-OCT-1997; 97JP-0285280.
 XX (FUJI) FUJISAWA PHARM CO LTD.
 XX Ishii Y, Noguchi Y, Saito Y, Soeda S, Yoshikawa K;
 XX WPI; 1999-302741/25.
 XX N-PSDB; AAX57909.
 XX Gene group for D-sorbitol dehydrogenase, useful for simple
 PT large-scale production of L-sorbose or 2-keto-L-gulononic acid as
 PT precursor for L-ascorbic acid
 XX
 XX Claim 7; Page 52-54; 83pp; Japanese.
 XX This sequence represents the D-sorbitol dehydrogenase of the
 CC invention. Cells transformed with a vector containing DNA encoding
 CC the dehydrogenase can be used to produce L-sorbose or 2-keto-L-gulononic
 CC acid as precursor for simple large-scale L-ascorbic acid production.
 XX
 XX Sequence 544 AA;
 SQ

Query Match 6.1%; Score 200; DB 20; Length 544;
 Best Local Similarity 20.2%; Pred. No. 1.7e-09;
 Matches 134; Conservative 76; Mismatches 220; Indels 232; Gaps 27;
 QY 37 DVFIAGSGPIGATYAKLCVZEAGLVRVYVMEIGAAADSFYVNAEETAVYYPVGHKKNEIE 96
 Db 9 dvvlgsgvgasanelaraglsivleagp-----rid 43
 QY 97 FKQDIDRF-VNVIKGALQOVSVPRNQNVPTLPDGAWSAPPSSAISNG-----KNPQRE 151
 Db 44 rghilenfrttenkgvqlypppv-----pwamhpbpdqgspngylhttgpdgae 92
 QY 152 FENLSAEAVTRGVGGSTHWCTSTPRHP---PMESLPGIGRPKLSNDPAEDDEKWEELY 208
 Db 93 yq-----qgylrvvvggtwnwagcawrylpdsfelhsrygvgr-----dwaikyddlpefy 143
 QY 209 SEAERLIGTSTKEFDESIRHTLVLSLQDAYKDRQRIF--RPLPLACHRLKNAPEYVEWH 266
 Db 144 yqaevmmgvagpamd-----vdlgsprshnypmkvpls-----178
 QY 267 SAENLFHSIYNDDOKKLETLTLNHR-----CTRLAL 298
 Db 179 -----ygadqfrklkhektnyrvvhepqarntprpydkrptcegnnncmpicpigm 229
 QY 299 TGG-----YEKKIGAAAEVRNLLATRNPSQSOLDYSIM-----AKVYVLASG 338
 Db 230 yngihsvnhaeagariipnavvyrletdasnkxkvvvnyydpdknshrvgtgkffvvaah 289
 QY 339 AI-----GNPQILYNSG-----FSLQVT--PRNDSLIPNLGRYITEQP 375
 Db 290 clesakillllsaddknpgrgianssdqvgvrmmdhtgqlsfmgndslwpggrpllt---346
 QY 376 MAFQIVLROEFVDSVRDDPYGLPWKE--AVAQHIKAKNETDALPIPRDPEQVTPFT 433
 Db 347 -----slidsfrdg---pwsergaylvhmvd-----dnqvdfatgl-380
 QY 434 EEHPWHTQIHRDAFSYGAVGPEVDSRVIVD---LRWF---GATDPEANNLLVFQNDVQ 485
 Db 381 -----aiaqgyvgeleedqirygsshavrlfshnegiadpd--nrlslskthk 426
 QY 486 DGYSMPOPTFRYR---PSTASNVRAKMMADMCEVASNLG-----GYLPSPQPMDPGL 537
 Db 427 dvlgiphevyvki1peytkvscdhtkel.fkelmalmsgtdpqwkgyfp-----475
 QY 538 ALHLACTTRIGFDKATVADNNSLVWDFANLYVAGNGTIRTGEGENPTLTSMCHAISAR 597
 Db 476 qchpsgstimgtdptnsvvdgecrthdhenlfvarsavfsvgtgnitltigalirvaa 535
 QY 598 SI 599
 Db 536 SI 537

RESULT 10
 ID AAW87541
 XX AAW87541 standard; Peptide; 51 AA.
 AC AAW87541;
 XX
 DT 26-FEB-1999 (first entry)
 XX Peptide derived from an antitumour protein.
 DE Antitumour protein; Tricholoma matsutake.
 XX
 XX Tricholoma matsutake.
 XX JP10313876-A.
 XX 02-DEC-1998.
 XX

```

XX OS Lyophyllum shimeji.
XX PN WO200121657-A1.
XX PD 29-MAR-2001.
XX PF 20-SEP-2000; 2000WO-JP06404.
XX PR 21-SEP-1999; 99JP-0267238.
XX PA (NISR) JAPAN TOBACCO INC.
XX PA (NORQ) SOC TECHNO-INNOVATION AGRIC FORESTY & FI.
XX PI Takakura Y, Kuwata S, Inoue Y;
XX DR WPI; 2001-281598/29.
XX CC Antibacterial protein and encoded gene isolated from Lyophyllum
XX PT shimeji, with activity against plant pathogenic bacteria, applicable in
XX PT agriculture e.g. rice cultivation at low concentration, produced at low
XX PT cost on large scale
XX PS Example 2; Page 24; 52pp; Japanese.
XX CC The present sequence is part of an antibacterial protein from the fungus
XX CC Lyophyllum shimeji. The protein was obtained from a fraction prepared
XX CC by extracting Lyophyllum shimeji with water and subjecting the extract
XX CC to ammonium sulphate precipitation. The protein inhibits the growth
XX CC of the plant pathogenic bacteria Pyricularia oryzae and Rhizoctonia
XX CC solani at a relatively low concentration. P. oryzae and R. solani are
XX CC causative of the two major diseases of rice. The protein contains
XX CC components of 70 kDa and 65 kDa, as determined by SDS-PAGE.
XX CC The antibacterial protein can be produced at low cost on a large scale.
XX SQ Sequence 34 AA;

Query Match 5.1%; Score 168; DB 22; Length 34;
Best Local Similarity 100.0%; Pred. No. 1.5e-08;
Matches 34; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 211 AERLIGTSTKEFDESIRHTLVRLSLQDAYKDROR 244
   |||||||
DB 1 aerlgtstkefdesirhtlvrlslqdaykdrqr 34

RESULT 15
AAB97031
ID AAB97031 standard; Peptide; 30 AA.
XX AC AAB97031;
XX DT 20-JUL-2001 (first entry)
XX DE Lyophyllum shimeji antibacterial protein fragment #1.
XX KW Fungus; antibacterial; antibiotic; plant pathogen; bacterial infection;
XX KW Pyricularia oryzae; Rhizoctonia solani; rice pathogen.
XX OS Lyophyllum shimeji.
XX PN WO200121657-A1.
XX PD 29-MAR-2001.
XX PF 20-SEP-2000; 2000WO-JP06404.
XX PR 21-SEP-1999; 99JP-0267238.
XX PA (NISR) JAPAN TOBACCO INC.
XX PA (NORQ) SOC TECHNO-INNOVATION AGRIC FORESTY & FI.

```

```

PI Takakura Y, Kuwata S, Inoue Y;
XX WPI; 2001-281598/29.
XX DR N-PSDB; AAF99980.
XX PT Antibacterial protein and encoded gene isolated from Lyophyllum
XX PT shimeji, with activity against plant pathogenic bacteria, applicable in
XX PT agriculture e.g. rice cultivation at low concentration, produced at low
XX PT cost on large scale
XX PS Claim 2; Page 32; 52pp; Japanese.
XX CC The present sequence is part of an antibacterial protein from the fungus
XX CC Lyophyllum shimeji. The protein was obtained from a fraction prepared
XX CC by extracting Lyophyllum shimeji with water and subjecting the extract
XX CC to ammonium sulphate precipitation. The protein inhibits the growth
XX CC of the plant pathogenic bacteria Pyricularia oryzae and Rhizoctonia
XX CC solani at a relatively low concentration. P. oryzae and R. solani are
XX CC causative of the two major diseases of rice. The protein contains
XX CC components of 70 kDa and 65 kDa, as determined by SDS-PAGE.
XX CC The antibacterial protein can be produced at low cost on a large scale.
XX SQ Sequence 30 AA;

Query Match 4.9%; Score 162; DB 22; Length 30;
Best Local Similarity 100.0%; Pred. No. 4.4e-08;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 76 NAEGETAVPYVPYGVYKKNETEFQKIDRFV 105
   |||||||
DB 1 naeegtavpyvpvygvynkneitefkdirdfv 30

```

Search completed: September 26, 2002, 11:06:56
Job time: 224 sec

DT 11-DEC-2000 (first entry)
 XX T. matsutake pyranose oxidase protein N-terminal fragment #2.
 DE Pyranose oxidase; matsutake mushroom; antibacterial; sugar metabolism;
 XX diabetes mellitus.
 KW Tricholoma matsutake.
 XX JP2000175698-A.
 XX 27-JUN-2000.
 XX 16-DEC-1998; 98JP-0357423.
 PF 16-DEC-1998; 98JP-0357423.
 PR (NLSB) JAPAN TOBACCO INC.
 XX WPI; 2000-478485/42.
 DR A reagent containing pyranose oxidase for the determination of pyranose
 PT Claim 3; Page 19; 23pp; Japanese.
 XX This invention describes a novel reagent containing pyranose oxidase for
 CC the determination of pyranose which can be prepared from a fraction
 CC precipitated from an aqueous extract of matsutake mushroom by ammonium
 CC sulfate precipitation and has an antibacterial activity against at least
 CC Pyricularia oryzae or Thanatephorus cucumeris and has a M.W. of ca.
 CC 20 kD and ca. 15 kD by SDS-PAGE and maintains the above antibacterial
 CC activity by being heated at 60 degrees C for 10 minutes in a neutral
 CC aqueous solution and in which the above antibacterial activity is
 CC inactivated by being heated at 80 degrees C for 10 minutes in a neutral
 CC aqueous solution. The invention also describes a method for the
 CC determination of pyranose in a sample in which the above reagent for the
 CC determination of pyranose is reacted with pyranose in the sample and the
 CC hydrogen peroxide formed is reacted with an enzyme to develop a color, a
 CC method for diagnosing a disease accompanied by abnormality in sugar
 CC metabolism in which the above reagent for the determination of pyranose
 CC is reacted with pyranose in the sample and the hydrogen peroxide formed
 CC is reacted with an enzyme to develop a color, and a kit for pyranose
 CC analysis or the diagnosis of a disease accompanied by abnormality in
 CC sugar metabolism. The pyranose oxidase can be used as a diagnostic agent
 CC for diabetes mellitus. This sequence represents a Tricholoma matsutake
 CC pyranose oxidase protein N-terminal fragment which is described in the
 CC method of the invention.
 XX Sequence 67 AA;
 SQ

Query Match 5.4%; Score 177; DB 21; Length 67;
 Best Local Similarity 57.4%; Pred. No. 6.8e-09;
 Matches 35; Conservative 6; Mismatches 20; Indels 0; Gaps 0;

QY 484 VODGYSMPQPTFRYPSTASNVRRKMMADMCEVASNLGGYLPTSPQFMDPGLALHLAG 543
 DB 7 vcdtympqptfhvkrtnadgrdgrmndmtanvnmlyggyipgysppqfmapgvlvhitg 66
 QY 544 T 544
 DB 67 t 67

RESULT 13
 ID AAY81951 standard; peptide; 67 AA.
 XX AAY81951;
 AC AAY81951;
 XX 07-JUL-2000 (first entry)
 DT

XX N-terminal fragment of Trichoderma derived antifungal protein.
 DE Antifungal protein; Pyricularia oryzae; Rhizoctonia solani; mushroom;
 XX growth inhibitor; plant pathogenic fungi; antibacterial agent;
 KW N-terminal fragment.
 XX Trichoderma matsutake.
 OS WO200014242-A1.
 XX 16-MAR-2000.
 PD 19-AUG-1999; 99WO-JP04441.
 XX 08-SEP-1998; 98JP-0270606.
 XX (NLSB) JAPAN TOBACCO INC.
 PA (NORQ) SOC TECHNO-INNOVATION AGRIC FORESTY & FI.
 XX Takakura Y, Kuwata S, Ohta S;
 PI WPI; 2000-256990/22.
 DR Mushroom-derived antibacterial protein against plant pathogenic fungi
 XX of rice, with activity and thermal stability, obtainable cheaply on
 PT large scale, useful in agriculture
 PS Claim 2; Page 27; 52pp; Japanese.
 XX This sequence represents an N-terminal fragment of the protein of the
 CC invention. The protein is an antibacterial protein with activity against
 CC at least Pyricularia oryzae and Rhizoctonia solani. It is obtained from a
 CC fraction of an aqueous extract of a mushroom precipitated by the ammonium
 CC sulphate precipitation method. The protein has a molecular weight of
 CC about 210 kD as determined by the gel filtration method, includes
 CC components of about 15 kD and 50 kD in SDS-PAGE, and is stable to
 CC heating in an aqueous neutral solution at 60 degrees C for 10 minutes but
 CC with loss of antibacterial activity after heating in the solution at 80
 CC degrees C for 10 minutes. The protein is used for inhibiting the growth
 CC of plant pathogenic fungi e.g. Pyricularia oryzae and Rhizoctonia solani.
 CC It is useful in treating rice plants, and is applicable in agriculture as
 CC an antibacterial agent. The protein has activity at relatively low
 CC concentrations, and can be produced at low cost on large scale.
 XX Sequence 67 AA;
 SQ

Query Match 5.4%; Score 177; DB 21; Length 67;
 Best Local Similarity 57.4%; Pred. No. 6.8e-09;
 Matches 35; Conservative 6; Mismatches 20; Indels 0; Gaps 0;

QY 484 VODGYSMPQPTFRYPSTASNVRRKMMADMCEVASNLGGYLPTSPQFMDPGLALHLAG 543
 DB 7 vcdtympqptfhvkrtnadgrdgrmndmtanvnmlyggyipgysppqfmapgvlvhitg 66
 QY 544 T 544
 DB 67 t 67

RESULT 14
 ID AAB97034 standard; Peptide; 34 AA.
 XX AAB97034;
 AC AAB97034;
 XX 20-JUL-2001 (first entry)
 DT Lyophyllum shimeji antibacterial protein fragment #4.
 DE Fungus; antibacterial; antibiotic; plant pathogen; bacterial infection;
 KW Pyricularia oryzae; Rhizoctonia solani; rice pathogen.
 XX

```

PF 13-FEB-1998; 98JP-0031452.
XX
XX 13-FEB-1997; 97JP-0029275.
XX
XX (MOMO-) MOMOYA KK.
PA (NORQ ) NORINSUISANSHO SHOKUJIN SOGO.
XX
XX WPI; 1999-074153/07.
DR
XX
XX An anti-tumour protein - prepared by culture of host cell
PT transformed by vector containing base coding sequence
PT
XX
XX Claim 5; Page 12; 15pp; Japanese.
PS
XX
XX The present sequence represents a peptide derived from an antitumour
CC protein. Nucleic acid encoding the antitumour protein is isolated
CC from Tricholoma matsutake.
XX
XX Sequence 51 AA;
XX
XX
XX Query Match 5.88; Score 192; DB 20; Length 51;
Best Local Similarity 62.7%; Pred. No. 1.6e-10;
Matches 32; Conservative 6; Mismatches 13; Indels 0; Gaps
XX
XX QY 405 VAOIHKNPDTALPIPRDPDPQVTPTEHPWHQIHRDAFSYGAVGPE 455
DB 1 veahkhhpdvlpipfhepeqqmipytsdfpwhvqhhrdafsygdvgpk 51
XX
XX
XX RESULT 11
XX AAY05719
XX ID AAY05719 standard; Protein; 615 AA.
XX
XX AC AAY05719;
XX
XX DT 19-JUL-1999 (first entry)
XX
XX DE Glucuronate dehydrogenase subunit I.
XX
XX KW Glucuronate dehydrogenase; membrane-bound; 2-keto-gluconate;
XX vitamin C; ascorbic acid.
XX
XX OS Erwinia cyripedii.
XX
XX FH Key Location/Qualifiers
XX FT Peptide 1..22
XX FT Protein /note="signal peptide"
XX FT Protein 23..615
XX FT Binding-site /note="mature protein"
XX FT /note="FAD binding motif"
XX
XX WO9915673-A1.
XX
XX PD 01-APR-1999.
XX
XX PF 25-SEP-1998; 98WO-KR00296.
XX
XX PR 25-SEP-1997; 97KR-0048802.
XX
XX PA (KOAD ) KOREA ADV INST SCI & TECHNOLOGY.
XX PA (KOMI-) KOREA MICROBIAL TECHNOLOGY INC.
XX
XX PI Pan JG, Yum DY;
XX
XX WPI; 1999-254718/21.
XX N-PSDB; AAX25345.
XX
XX New isolated gluconate dehydrogenase
XX
XX Claim 13; Page 22-23; 42pp; English.
XX

```


CC Pyricularia oryzae or Thanatephorus cucumeris and has a M.W. of ca.
 CC 210 kD by gel filtration and shows the presence of components of ca.
 CC 50 kD and ca. 15 kD by SDS-PAGE and maintains the above antibacterial
 CC activity by being heated at 60 degrees C for 10 minutes in a neutral
 CC aqueous solution and in which the above antibacterial activity is
 CC inactivated by being heated at 80 degrees C for 10 minutes in a neutral
 CC aqueous solution. The invention also describes a method for the
 CC determination of pyranose in a sample in which the above reagent for the
 CC determination of pyranose is reacted with pyranose in the sample and the
 CC hydrogen peroxide formed is reacted with an enzyme to develop a color, a
 CC method for diagnosing a disease accompanied by abnormality in sugar
 CC metabolism in which the above reagent for the determination of pyranose
 CC is reacted with pyranose in the sample and the hydrogen peroxide formed
 CC is reacted with an enzyme to develop a color, and a kit for pyranose
 CC analysis of the diagnosis of a disease accompanied by abnormality in
 CC sugar metabolism. The pyranose oxidase can be used as a diagnostic agent
 CC for diabetes mellitus. This sequence represents the Trichoderma matsutake
 CC pyranose oxidase protein which is described in the method of the
 CC invention.
 XX
 SQ Sequence 564 AA;

Query Match 31.8%; Score 1044; DB 21; Length 564;
 Best Local Similarity 40.1%; Pred. No. 7.7e-89;
 Matches 243; Conservative 94; Mismatches 197; Indels 72; Gaps 17;

QY 17 QINGQIPKNAIHTYNDGV-----DVFAGSGPIGATYAKLCVE--AGLRVWVVEIGAAD 70
 DB 9 kindllqrsqgdltssqdeivhytdvfiagsgpiactyarihiidntsttkvymaiegsqd 68
 QY 71 SFYAVNAEETATVPYVPGYHKKEIEFQKIDRFVNVIKGALQOVSPVRNQNVPTLDGP 130
 DB 69 n-----pvi-gahnrnsikfkqdktdkfvnlngalqpsispsdyqptlava 115
 QY 131 AWSAPPSSA-----ISNGKNPHOREPENLSAEAVTRGVGMSTHTCTSTPRIHPMESLP 186
 DB 116 aw-appidpaegqlvmghnpnqaglnlpgsavtrtvggmawhwtacacphtphd----- 168
 QY 187 GICRPLKSNPDADDKWNELYSAEERLIGTSYKEFDESIRHILVLSLODAYKDRORIF 246
 DB 169 -----eervnnpv-dkqefdalleraktlinvhsdqydsirqvketlqgtl-dasrgv 222
 QY 247 RPLFLACHRLKNAPEYVEVHSAENLFHSIYNDKQKFLTLNHRCTRLALTGGEKKI 306
 DB 223 ttllplovertdnpiyvtgtdad-----tvlgdvpkspfrvlvteetrkfvsetnptqv 278
 QY 307 GAEEVRLNLTARPPSSOLDYIMAKVTVLASGAIGNPOILYNSGFSGLQVTPRNDSLIPN 366
 DB 279 vaallrnl-----ntsn--delvvaqsvfiacgavctpqilwns-----nirph-----a 321
 QY 367 LGRVITEOPNAFCQIVLROEFVDSRDDPYGLPMWKEAVAQHTAKNPTDALPIFRDPEP 426
 DB 322 lgrviseqsmfqcqvlkrsivdsiaedpr-----faakveahkkhdpdvlpihphepe 377
 QY 427 QVTPTEEPHPTQIHRDAFSGAVGPEVDSRVIVDLRFWFGATDPEANLLVE----- 480
 DB 378 qmlyptdsdfpwhvqhr--yafgdvgpkadprvvdrlfrfgksdlveearvtfgpnkl 435
 QY 481 ---QNDVQDSYMPQPTFRYRSTASVNRKAKMADCEVASNIGGLYPTSPPOFMDPGL 537
 DB 436 rdweagvtdtygmpqptfthvkrtnadgdrqrmndntnvanllggyipgsypqfmapgl 495
 QY 538 ALHLAGTTRIGFDPKATVADNNSLWDFANLYVAGNCTIRTGGENPTLSMCHAIKSAR 597
 DB 496 aqhtgttrigtqdsvadqtskvhnfnlhwggngcipdatacnptrtsvayalkgae 555
 QY 598 SIINTL 603
 DB 556 avvsyl 561

RESULT 8

AAV81952

ID AAY81952 standard; Protein; 564 AA.

XX AAY81952;

AC AAY81952;

XX 07-JUL-2000 (first entry)

DE Trichoderma derived antifungal protein sequence.

XX Antifungal protein; Pyricularia oryzae; Rhizoctonia solani; mushroom;

KW growth inhibitor; plant pathogenic fungi; antibacterial agent;

KW N-terminal fragment.

XX Trichoderma matsutake.

OS WO200014242-A1.

XX 16-MAR-2000.

XX 19-AUG-1999; 99WO-JP04441.

XX 08-SEP-1998; 98JP-0270606.

XX (NISR) JAPAN TOBACCO INC.

PA (NORQ) SOC TECHNO-INNOVATION AGRIC FORESTY & FI.

XX Takakura Y, Kuwata S, Ohta S;

PI WPI; 2000-256990/22.

DR N-PSDB; AAA07403.

XX Mushroom-derived antibacterial protein against plant pathogenic fungi

PT of rice, with activity and thermal stability, obtainable cheaply on

PT large scale, useful in agriculture

XX Claim 3; Page 41-45; 52pp; Japanese.

CC This sequence is the Trichoderma matsutake antibacterial protein of

CC the invention. The protein has activity against at least Pyricularia

CC oryzae and Rhizoctonia solani, and is obtained from a fraction of an

CC aqueous extract of a mushroom precipitated by the ammonium sulphate

CC precipitation method. The protein has a molecular weight of about 210 kD

CC as determined by the gel filtration method, includes components of about

CC 15 kD and 50 kD in SDS-PAGE, and is stable to heating in an aqueous

CC neutral solution at 60 degrees C for 10 minutes but with loss of

CC antibacterial activity after heating in the solution at 80 degrees C for

CC 10 minutes. The protein is used for inhibiting the growth of plant

CC pathogenic fungi e.g. Pyricularia oryzae and Rhizoctonia solani.

CC It is useful in treating rice plants, and is applicable in agriculture as

CC an antibacterial agent. The protein has activity at relatively low

CC concentrations, and can be produced at low cost on large scale.

XX Sequence 564 AA;

SQ

Query Match 31.8%; Score 1044; DB 21; Length 564;

Best Local Similarity 40.1%; Pred. No. 7.7e-89;

Matches 243; Conservative 94; Mismatches 197; Indels 72; Gaps 17;

QY 17 QINGQIPKNAIHTYNDGV-----DVFAGSGPIGATYAKLCVE--AGLRVWVVEIGAAD 70

DB 9 kindllqrsqgdltssqdeivhytdvfiagsgpiactyarihiidntsttkvymaiegsqd 68

QY 71 SFYAVNAEETATVPYVPGYHKKEIEFQKIDRFVNVIKGALQOVSPVRNQNVPTLDGP 130

DB 69 n-----pvi-gahnrnsikfkqdktdkfvnlngalqpsispsdyqptlava 115

QY 131 AWSAPPSSA-----ISNGKNPHOREPENLSAEAVTRGVGMSTHTCTSTPRIHPMESLP 186

DB 116 aw-appidpaegqlvmghnpnqaglnlpgsavtrtvggmawhwtacacphtphd----- 168

QY 187 GICRPLKSNPDADDKWNELYSAEERLIGTSYKEFDESIRHILVLSLODAYKDRORIF 246

DB 169 -----eervnnpv-dkqefdalleraktlinvhsdqydsirqvketlqgtl-dasrgv 222

QY 247 RPLFLACHRLKNAPEYVEVHSAENLFHSIYNDKQKFLTLNHRCTRLALTGGEKKI 306

DB 223 ttllplovertdnpiyvtgtdad-----tvlgdvpkspfrvlvteetrkfvsetnptqv 278

QY 307 GAEEVRLNLTARPPSSOLDYIMAKVTVLASGAIGNPOILYNSGFSGLQVTPRNDSLIPN 366

DB 279 vaallrnl-----ntsn--delvvaqsvfiacgavctpqilwns-----nirph-----a 321

QY 367 LGRVITEOPNAFCQIVLROEFVDSRDDPYGLPMWKEAVAQHTAKNPTDALPIFRDPEP 426

DB 322 lgrviseqsmfqcqvlkrsivdsiaedpr-----faakveahkkhdpdvlpihphepe 377

QY 427 QVTPTEEPHPTQIHRDAFSGAVGPEVDSRVIVDLRFWFGATDPEANLLVE----- 480

DB 378 qmlyptdsdfpwhvqhr--yafgdvgpkadprvvdrlfrfgksdlveearvtfgpnkl 435

QY 481 ---QNDVQDSYMPQPTFRYRSTASVNRKAKMADCEVASNIGGLYPTSPPOFMDPGL 537

DB 436 rdweagvtdtygmpqptfthvkrtnadgdrqrmndntnvanllggyipgsypqfmapgl 495

QY 538 ALHLAGTTRIGFDPKATVADNNSLWDFANLYVAGNCTIRTGGENPTLSMCHAIKSAR 597

DB 496 aqhtgttrigtqdsvadqtskvhnfnlhwggngcipdatacnptrtsvayalkgae 555

PT cost on large scale
XX Claim 3; Page 43-45; 52pp; Japanese.
XX
XX The present sequence is an antibacterial protein from the fungus
CC Lyophyllum shimeji. The protein was obtained from a fraction prepared
CC by extracting Lyophyllum shimeji with water and subjecting the extract
CC to ammonium sulphate precipitation. The protein inhibits the growth
CC of the plant pathogenic bacteria Pyricularia oryzae and Rhizoctonia
CC solani at a relatively low concentration. P. oryzae and R. solani are
CC causative of the two major diseases of rice. The protein contains
CC components of 70 kDa and 65 kDa, as determined by SDS-PAGE.
CC The antibacterial protein can be produced at low cost on a large scale.
XX
XX Sequence 618 AA;

Query Match 100.0%; Score 3284; DB 22; Length 618;
Best Local Similarity 100.0%; Pred. No. 2.1e-299;
Matches 618; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MSLTEQMLRDYPRSMQINGOIPKNAHETGNDGVDFIAGSGPIGATYAKLCVEAGLR 60
DB 1 mslstegmlrdyprsmqmgdipknahtyngdgvdfiagsgpigtatyaklcveaglr 60
QY 61 VYVVEIGAADSFYAVNAEETAVPYVPGYHKKNEIEFQKIDIDRFVNVIKGALQOVSPVR 120
DB 61 vvmveigaadsfyavnaeegtavpyvpgyhhkneiefqkldidrfvnnvlgkalgqvsvpr 120
QY 121 NVNPTLDPGAWSAPPSSAISNGKNPHOREFENLSABAVTRGVGMSHTWCSTPRIHP 180
DB 121 nqnvtldpgawsappssalsngknphqrefenlsaaavtrvggmshwtwcstprihp 180
QY 181 PMESLPGIGRKLNDPAEDDKENELYSERLIGSTKEFDESIRTLVLSLQDAYK 240
DB 181 pmeslpgigrklndpaeddkewnelysesarligstkefdesirhtlvlsrlqdayk 240
QY 241 DRQIFRPLPLACHRLKNAPYVWHSANLFIHSYNDKOKKLTLLTNHRCRRLALTG 300
DB 241 drqifrlplachrlknapeyvwhsaenlfhsynddkkklftlltnhrcrtraltg 300
QY 301 GYEKIGAAEVRNLLATRNPSQSDSYIMAKVYVLASGAINQPILYNSGFGLOVTPRN 360
DB 301 gyeekigaaevrnllatrnpsqsdysimakvylvlasgaignpqillynsgfglqvtpn 360
QY 361 DSLIPNLGRYITEOPMAFCQVLROEFVDSVRDDPYGLPWKEAVAQHIANKPTDALPI 420
DB 361 dslipnlgrYTEOPMAFCQVLROEFVDSVRDDPYGLPWKEAVAQHIANKPTDALPI 420
QY 421 FRDPEQVPTTPEEHPWHTQIHRDAFSYGAVGPEVDSRVTVDLRFWFGATDPEANNLV 480
DB 421 frdpeqvpttteeHPWHTQIHRDAFSYGAVGPEVDSRVTVDLRFWFGATDPEANNLV 480
QY 481 QNVODGYSMPQPTFRPTASTNVRARKMADCEVASNIGGLYLPSPFPMDPGLALH 540
DB 481 qndvqdgysmpqptfyrptstasnvrarkmadcevasnlgglylpsppfmdpglalh 540
QY 541 LAGTRTGFQKATVANNLSLWDFANLYVAGNCTIRTFGENTLTSMCHAIKSARSII 600
DB 541 lagtrtgrfkatvannslswdfanlyvagnctirtfgenltltsmchaiksarsii 600
QY 601 NTLKGGTDGKNTGEHRLN 618
DB 601 ntlkgtgdgkntgehrnl 618

RESULT 2
ID AAW94308
XX AAW94308 standard; protein; 623 AA.
XX AAW94308;
XX
DT 08-APR-1999 (first entry)

XX Coriolus versicolor pyranose oxidase.
DE
XX Coriolus versicolor; modified; pyranose oxidase; heat stability;
KW pH stability; glucose; glucosone.
XX Coriolus versicolor.
OS
XX JPI1009271-A.
PN
XX 19-JAN-1999.
PD
XX 24-JUN-1997; 97JP-0181865.
PF
XX 24-JUN-1997; 97JP-0181865.
PR
XX (KIKK) KIKKOMAN CORP.
PA
XX WPI; 1999-145891/13.
DR
XX New modified pyranose oxidase gene and protein - useful for
PT efficient recombinant production of pyranose oxidase with excellent
PT Km, heat and pH stability
XX
PS Claim 1; Page 7-9; 10pp; Japanese.
XX
CC The present sequence represents pyranose oxidase derived from Coriolus
CC versicolor, which can be modified by adding, deleting or replacing at
CC least one amino acid (preferably where the amino acid at position 542
CC is replaced) and still retains pyranose oxidase activity. The modified
CC pyranose oxidase with the amino acid at position 542 replaced, has the
CC following properties: (a) Action: It oxidises glucose to glucosone;
CC (b) Substrate specificity: It reacts specifically with glucose and also
CC reacts with galactose, L-sorbose, D-xylitol and 1,5-anhydro-D-glucitol;
CC (c) Stable pH: 3.5 to 11.0 at 50 degrees Celsius for 30 minutes; (d)
CC Optimum pH: Near 6.5; (e) Optimum temperature: Near 55 degrees Celsius;
CC and (f) Temperature stability: Stable up to 55 degrees Celsius. The
CC present invention also describes a method for the preparation of a
CC modified pyranose oxidase. The method efficiently prepares a modified
CC pyranose oxidase with excellent Km value, heat stability and
XX pH stability.
SQ Sequence 623 AA;

Query Match 37.5%; Score 1232.5; DB 20; Length 623;
Best Local Similarity 45.0%; Pred. No. 1.7e-106;
Matches 273; Conservative 73; Mismatches 176; Indels 85; Gaps 16;
QY 37 DVFIAGSGPIGATYAKLCVEAGLRVWVEIGAADSFYAVNAEETAVPYVPGYHKKNEIE 96
DB 48 dvlvlgsgpigtatyaklcveaglvwvveigaadsvfavnadeegtavpyvpgyhhkntve 95
QY 97 FQKIDIRFVNVYKALQOVSVPRNVQNVPTLDPGAWSAPPSSAISNGKNPHOREFENLS 156
DB 96 yqkndkrfvnykqglmsvsvpntlvldtltspstsqda--ssffvngnspeqdpnlsls 153
QY 157 AEATRVGGVSTHTWCTSTPRIHPMESLPGIGRKLNDPAE--DDKEWNELYSEARLI 215
DB 154 gqavtrvggmshwtcatprfdreq-----rpllvkddqddaedwrdlyckaesyf 206
QY 216 GTSTKEFDESIRHTLVLSRLQDAYKDRQIFRPLPLACHRLKNAPYVWHSANLFI--H 273
DB 207 ktgtdqfkesirhnlvinklaeeykg-qrdfqqlplaatr--rsptfveassantvdlq 263
QY 274 SIYNDKOKKLTLLTNHRCRRLALTGGYEKKIGAAEVRNLLATRNPSQSDSY----- 327
DB 264 nrpntdapnerfnlpavacerv-----vln-----tsnseieslhldli 304
QY 328 -----IMAKVYVLASGAINQPILYNSGFGSL-QVTPRN-DSLIPNLGRYITEOPMAFC 379
DB 305 sgdrfeikadvitavghnaqilvnsqfqlgrpdpnppqilpalsyiteqsivfc 364

Db 462 RNOQCVESNOAMPVSRPPSSVLTDD--SYTSSIP---ANNVLDRRKTCRLVELISDQR 516
 QY 393 -DDPYGLPMWKEVAQHIANKPTDALP 418
 Db 517 TDDPELLEFY-HMVKEVRARYPHDDAP 542

RESULT 14

US-09-023-731-6
 ; Sequence 6, Application US/09023731
 ; Patent No. 6291648
 ; GENERAL INFORMATION:
 ; APPLICANT: Kawamura, Yukio; Morita,
 ; APPLICANT: Akhito; Izumo, Koji.; Saka, Tomohide.
 ; TITLE OF INVENTION: ANTITUMOR PROTEIN AND
 ; TITLE OF INVENTION: GENE ENCODING SAME
 ; NUMBER OF SEQUENCES: 20
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: MORGAN & FINNEGAN, L.L.P.
 ; STREET: 345 PARK AVENUE
 ; CITY: NEW YORK
 ; STATE: NEW YORK
 ; COUNTRY: USA
 ; ZIP: 10154
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: FLOPPY DISK
 ; COMPUTER: IBM PC COMPATIBLE
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: MICROSOFT WORD 97
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/09/023,731
 ; FILING DATE: 13-FEB-1998
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: JP 29275/1997
 ; FILING DATE: 13-FEB-1997
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: KENNETH H. SONNENFELD
 ; REGISTRATION NUMBER: 33,285
 ; REFERENCE/DOCKET NUMBER: 3479-4000
 ; TELEPHONE: (212) 758-4800
 ; TELEFAX: (212) 751-6849
 ; TELEX: 421792
 ; INFORMATION FOR SEQ ID NO: 6:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 61
 ; TYPE: amino acid
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: peptide
 ; US-09-023-731-6

Query Match 3.1%; Score 102.5; DB 4; Length 61;
 Best Local Similarity 43.5%; Pred. No. 0.0027;
 Matches 27; Conservative 6; Mismatches 24; Indels 5; Gaps 2;
 QY 104 FNVVIGALQOVSVFVRNQNVPVLDPCGAWSPGSSA-----ISNGKNPHOREFENLSAE 159
 Db 1 FVNIINGALQPIISPSDTYQPTLAVAAW-APPIDPAEGOLVINGHNPNOEAGINLP 59
 ; QY 160 VT 161
 ; QY 60 VT 61

RESULT 15

US-08-549-515-5
 ; Sequence 5, Application US/08549515
 ; Patent No. 6034123
 ; GENERAL INFORMATION:
 ; APPLICANT: Loosmore, Sheena M
 ; APPLICANT: Klein, Michel H
 ; TITLE OF INVENTION: Haemophilus Influenzae

; TITLE OF INVENTION: Dimethylsulphoxide Reductase Enzyme
 ; NUMBER OF SEQUENCES: 12
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Sim & McBurney
 ; STREET: Suite 701, 330 Unviersity Avenue
 ; CITY: Toronto
 ; STATE: Ontario
 ; COUNTRY: Canada
 ; ZIP: M5G 1R7
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patentin Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/549,515
 ; FILING DATE: 27-OCT-1995
 ; CLASSIFICATION: 536
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Stewart, Michael I.
 ; REGISTRATION NUMBER: 24,973
 ; REFERENCE/DOCKET NUMBER: 1038-522
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (416) 595-1155
 ; TELEFAX: (416) 595-1163
 ; INFORMATION FOR SEQ ID NO: 5:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 806 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; US-08-549-515-5

Query Match 3.1%; Score 100.5; DB 3; Length 806;
 Best Local Similarity 18.6%; Pred. No. 0.42;
 Matches 69; Conservative 56; Mismatches 139; Indels 107; Gaps 17;
 QY 31 YGNDGVDVFIAGSGPIGATYAKLCVE-----56
 Db 230 FGNNPAETRMSSG---GLTY---CIEQAKARSNAKMIIDPRYNDTGAGREDEWIPIRPG 283
 QY 57 -----AGLVVAVVEIGAAD-----SFYAVNAEAGTAVPYVP--GYHKKNEIEFKDIDRF 104
 Db 284 TDAALVAALAYVMIQENLVDPQFLDKYCVGYDEKTLPADAPKNGHYKAYILGYGN--DGI 341
 QY 105 VNVIKGALQOVSVFVRNQNVPVLDPCGAWSPGSSAISNGKNPHOREFENLSAEAVTRGV 164
 Db 342 AKTPENAAKITGIPA--ERIKLAREIGSTKP--AFISQGWGQRRS-----NGELISR 393
 QY 165 -----GGMSTHTWCSTPRIHPMBESLPGIGRPKLSNDPAEDDKWENLYSEARLIGT 217
 Db 394 AMLPILTNVGIHGGNTGAR--ESAYSIFVRMPTLK-----NPMKASIPMELGT 441
 QY 218 STEFDESIRHTLVLSLODAYKDRQIFRPLPLACHRLKNAPEYVHSAENLFHSIYN 277
 Db 442 -----DAIIIRGT-EMTALTDIRGVKLSPPKIVWYNASNC--LINQHAQINRTHDILQ 493
 QY 278 DDQKKLFTLLTNHRCRTRALTGTGKKGAAAEVRNLATRNPSQ-----LDSYIMAKV 332
 Db 494 DDTQCEMIITIDNHMTS-----TAKYSIDILLPDCPTSEQMDALDAFVSNMA 540
 QY 333 YVLASGAIGNP 343
 Db 541 YVIFADQVIKP 551

Search completed: September 26, 2002, 11:07:19
 Job time: 217 sec

```

; SEQUENCE CHARACTERISTICS:
; LENGTH: 50
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-09-023-731-14

Query Match 3.5%; Score 115; DB 4; Length 50;
Best Local Similarity 53.5%; Pred. No. 9.7e-05;
Matches 23; Conservative 5; Mismatches 15; Indels

Oy 484 VODGYSMPQPTFRYRSTASNVARKMMADCEVASINLGGYLP 526
| | | | | | | | | | | | | | | | | | | | |
Db 8 VTDTYGMPQPTFRVHKRTNADGDRDQRMNDNTVNYANMLGGYLP 50

RESULT 11
US-09-023-731-15
; Sequence 15, Application US/09023731
; Patent No. 6291648
; GENERAL INFORMATION:
; APPLICANT: Kawamura, Yukio; Morita,
; APPLICANT: Akihiro; Izumo, Koji.; Saka, Tomohide.
; TITLE OF INVENTION: ANTITUMOR PROTEIN AND
; TITLE OF INVENTION: GENE ENCODING SAME
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORGAN & FINNEGAN, L.L.P.
; STREET: 345 PARK AVENUE
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: USA
; ZIP: 10154

```

```

MEDIUM TYPE: FLOPPY DISK
COMPUTER: IBM PC COMPATIBLE
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: MICROSOFT WORD 97
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/023,731
FILING DATE: 13-FEB-1998
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 29275/1997
FILING DATE: 13-FEB-1997
ATTORNEY/AGENT INFORMATION:
NAME: KENNETH H. SONNENFELD
REGISTRATION NUMBER: 33,285
REFERENCE/DOCKET NUMBER: 3479-4000
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 758-4800
TELEFAX: (212) 751-6849
TELEX: 421792
INFORMATION FOR SEQ ID NO: 15:
SEQUENCE CHARACTERISTICS:
LENGTH: 35
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-09-023-731-15

Query Match 3.3%; Score 108; DB 4; Length 35;
Best Local Similarity 66.7%; Pred. No. 0.00028;
Matches 22; Conservative 2; Mismatches 9; Indels

Qy 528 SPQFMDFGGLALHLAGTTRIGFGKATTVADNNS 560
|-----|
Db 2 SYPQFMAGPLVLHITGTTRIGTDQIQSVADPTS 34

RESULT 12
US-08-609-049A-12

```

RESULT 12
US-08-609-049A-12

ORGANISM: Erwinia cyripedii
US-09-297-937C-9

Query Match 5.8%; Score 191.5; DB 4; Length 615;
Best Local Similarity 21.1%; Pred. No. 9.5e-11;
Matches 134; Conservative 83; Mismatches 256; Indels 163; Gaps 32;

QY 36 VDFIAGSGPIGATYAKLCVEAGLRVVMVEIGA-----ADSFYAVNAEEGTAVVPGYH 90
DB 29 VDAVVGVGAGATMAKELTEAGLNVVALERGRHDPYDGYQSIDELT-----YN 81
QY 91 KKNIEFQKIDRFVNVKIGALQOVSPVRNQNVPTLDPGAWSPGSSAISNKNPHQR 150
DB 82 IRKRL-FQ-DLSKSTVIRHDASQTAVPYR----- 109
QY 151 EFENLSAEAVTRGVGGMSTHTWCSTPRHPPMESLPG-----IGRPKLSNDPAEDD--KE 203
DB 110 ---QLAAFLPGTGGAGLHWSGVHFRVDPVELNLRSHYEARYGKNTFPEGMTIODEGVS 166
QY 204 WNEL---YSEARLIGTSTKEFDESIRHTLVLSLQDAY--KDRQIRFRPLP-----LAC 253
DB 167 YNELEPFDOAEKVFGTSGSAW--TIKGMIGKEKGGNFYAPDRSSDF-PLPAOKRTYSA 223
QY 254 HRLKNAPEYVEWH-----SAENLFHSIYND----- 279
DB 224 QLFQAASVGYHYPDMPSANTSGPYNTYGAQMPCNFCGCGYACYMYSKASPNVNI 283
QY 280 ----KQKKLFTLLNHRCTRRLALGGYEKKIGA---AEVNLLATRNPSQ--LDSYIM 329
DB 284 LPALNQPKPELRNNAVLRVNLGDKKRATGVYLDGQGREVV---OPADLVLSAQF 340
QY 330 AKVYVLAGAIGNP-QILYNSGFGSLQVTPRNDLSIPLN-GRYTEOPM---AFQIVLR 384
DB 341 HNVHMLLSGIGQYPNITNEGVVGRNFAYQNI STLKALFDKNTTNPFFIGAGAGAVD 400
QY 385 QEFVDSVRDDPYGL-----PWWKEAVAQHIKNTDPLPIPRDPPEQVTPFTTEHPWHT 440
DB 401 DFNADNHDGPGYGVGSPFW-----VNOAGTKPVSGLTPKGTN-----WGS 444
QY 441 QIHRDAPSYGAVGEVDSRVIVDLRWFGATDPEANNLLVFQNDVQGYMSP--OPTERYR 498
DB 445 QWK-----AAVDYTHHHSMDAH--GAHOSYRANYLDLDPNKNYVGPQLLRMTDMQ 496
QY 499 PSTASNV-ARKMADCEV--ASN---LGGYLPSPQPMFPLG--ALHLAGTTRIGF 549
DB 497 D---NDIRMAQFVYGMKMKITEAMNPKMIIGG--AKGPGTHFTTIVYQTTTHMSGGAIMGE 551
QY 550 DKATTVADNNSLVWDFANLYVAGNGTIRTGEGENPT 585
DB 552 DPKTSVNRVYLSQWDPNVFVPGASAFPOGLGYNPT 587

RESULT 6

US-09-297-937C-13
Sequence 13, Application US/09297937C
Patent No. 6337199
GENERAL INFORMATION:
APPLICANT: YUN, Do Young
APPLICANT: PAN, Jae Gu
TITLE OF INVENTION: Membrane-Bound Gluconate Dehydrogenase, Gene Sequence
TITLE OF INVENTION: Encoding the Same and Production of 2-keto-D-Gluconate
FILE REFERENCE: P66159050
CURRENT APPLICATION NUMBER: US/09/297,937C
CURRENT FILING DATE: 1999-05-11
PRIOR APPLICATION NUMBER: PCT/KR98/00296
PRIOR FILING DATE: 1998-09-25
PRIOR APPLICATION NUMBER: KR 97-48802
PRIOR FILING DATE: 1997-09-25
NUMBER OF SEQ ID NOS: 13
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 13

LENGTH: 1276
TYPE: PRT
ORGANISM: Erwinia cyripedii
US-09-297-937C-13

Query Match 5.8%; Score 191.5; DB 4; Length 1276;
Best Local Similarity 21.1%; Pred. No. 3.4e-10;
Matches 134; Conservative 83; Mismatches 256; Indels 163; Gaps 32;

QY 36 VDFIAGSGPIGATYAKLCVEAGLRVVMVEIGA-----ADSFYAVNAEEGTAVVPGYH 90
DB 249 VDAVVGVGAGATMAKELTEAGLNVVALERGRHDPYDGYQSIDELT-----YN 301
QY 91 KKNIEFQKIDRFVNVKIGALQOVSPVRNQNVPTLDPGAWSPGSSAISNKNPHQR 150
DB 302 IRKRL-FQ-DLSKSTVIRHDASQTAVPYR----- 329
QY 151 EFENLSAEAVTRGVGGMSTHTWCSTPRHPPMESLPG-----IGRPKLSNDPAEDD--KE 203
DB 330 ---QLAAFLPGTGGAGLHWSGVHFRVDPVELNLRSHYEARYGKNTFPEGMTIODEGVS 386
QY 204 WNEL---YSEARLIGTSTKEFDESIRHTLVLSLQDAY--KDRQIRFRPLP-----LAC 253
DB 387 YNELEPFDOAEKVFGTSGSAW--TIKGMIGKEKGGNFYAPDRSSDF-PLPAOKRTYSA 443
QY 254 HRLKNAPEYVEWH-----SAENLFHSIYND----- 279
DB 444 QLFQAASVGYHYPDMPSANTSGPYNTYGAQMPCNFCGCGYACYMYSKASPNVNI 503
QY 280 ----KQKKLFTLLNHRCTRRLALGGYEKKIGA---AEVNLLATRNPSQ--LDSYIM 329
DB 504 LPALNQPKPELRNNAVLRVNLGDKKRATGVYLDGQGREVV---OPADLVLSAQF 560
QY 330 AKVYVLAGAIGNP-QILYNSGFGSLQVTPRNDLSIPLN-GRYTEOPM---AFQIVLR 384
DB 561 HNVHMLLSGIGQYPNITNEGVVGRNFAYQNI STLKALFDKNTTNPFFIGAGAGAVD 620
QY 385 QEFVDSVRDDPYGL-----PWWKEAVAQHIKNTDPLPIPRDPPEQVTPFTTEHPWHT 440
DB 621 DFNADNHDGPGYGVGSPFW-----VNOAGTKPVSGLTPKGTN-----WGS 664
QY 441 QIHRDAPSYGAVGEVDSRVIVDLRWFGATDPEANNLLVFQNDVQGYMSP--OPTERYR 498
DB 665 QWK-----AAVDYTHHHSMDAH--GAHOSYRANYLDLDPNKNYVGPQLLRMTDMQ 716
QY 499 PSTASNV-ARKMADCEV--ASN---LGGYLPSPQPMFPLG--ALHLAGTTRIGF 549
DB 717 D---NDIRMAQFVYGMKMKITEAMNPKMIIGG--AKGPGTHFTTIVYQTTTHMSGGAIMGE 771
QY 550 DKATTVADNNSLVWDFANLYVAGNGTIRTGEGENPT 585
DB 772 DPKTSVNRVYLSQWDPNVFVPGASAFPOGLGYNPT 807

RESULT 7

US-09-023-731-11
Sequence 11, Application US/09023731
Patent No. 6291648
GENERAL INFORMATION:
APPLICANT: Kawamura, Yukio; Morita,
APPLICANT: Akhino; Izumo, Koji.; Saka, Tomohide.
TITLE OF INVENTION: ANTITUMOR PROTEIN AND
TITLE OF INVENTION: GENE ENCODING SAME
NUMBER OF SEQUENCES: 20
CORRESPONDENCE ADDRESS:
ADDRESSEE: MORGAN & FINNEGAN, L.L.P.
STREET: 345 PARK AVENUE
CITY: NEW YORK
STATE: NEW YORK
COUNTRY: USA
ZIP: 10154
COMPUTER READABLE FORM:

Query Match 37.5%; Score 1232.5; DB 1; Length 623;
Best Local Similarity 45.0%; Pred. No. 9.8e-119;
Matches 273; Conservative 73; Mismatches 176; Indels 85; Gaps 16;

QY 37 DVFTAGSGPIGATYAKLCVEAGLRVVMVEIGAADSFYAVNAEEGTAVPYVPGYHKKNEIE 96
DB 48 DVFTAGSGPIGCTYARELVVVEAGLVVVEIGAADSFYAVNAEEGTAVPYVPGYHKKNEIE 95
QY 97 FOKDIDRFVNVKIGALQOVSVVVRNQVPTLDPGAWSPGSSAISNKNPHQREFENLS 156
DB 96 YQKNIDRFVNVQIGLMSVSVVNTLVITLSPYSWQA--SSEFVRNGSNPEQDPLRLNS 153
QY 157 AEATRVGGGSMTHWTCTPRIHPMESLPGIGRPKLSNDPAE--DDKEWNELYSEAEHLI 215
DB 154 GOAVTRVVGGMSTHWTCTATPRDEQ-----RPLLKDDODADDAEMDRLYKAESYF 206
QY 216 GTSKFEDESIRHTLVLSLQADAYKDRQIRFRPLPLACHRLKNAPEYVEVHSAENLF--H 273
DB 207 KTGTQDQKESIRHNLVNLKLAEEYKG--QRDQOIPLAATR--RSPTFEVWSSANTVFDLQ 263
QY 274 SIYNDKOKKLTLLTNHRCRLALTGGYKKGAAEVRNLLATRNPSQLDSY----- 327
DB 264 NRPTDAPNERENLFPVACERV-----VRN-----TSNSEIESLHIDLI 304
QY 328 ----IMAKVYVLAGAIGNPQILYNSGSGGL--QVTPRN--DSLIPNLGRYITEOPMAFC 379
DB 305 SGDRFEIKADVFLTAGAVHNAQLLVNSGFGQLGRPDANPPQLPSLGSYITEOSLVFC 364
QY 380 QIVLRQBFVDSVRDDPY--GLP-----WKKEAVAQIAHAKNPTDAL 417
DB 365 QTVMSTELIDSVKSDMIIRGNPNGLGYSVYTPGAETNKNKPDWNEKVKNNMQHQPDL 424
QY 418 PIPEDPEPQVTTPEEHPHWTQIHRDAFSYGAVGPEVDSRVIVDLRFWFGATDEANNL 477
DB 425 PIPEDPEPQVTTQFQSHPHWTQIHRDAFSYGAVQOQSDSLRILVDMFRFGTEPKEENK 484
QY 478 LVFONDQVQSGMPQPRYR--PSTASNVARKKMMADMCEVASNLGGYLPSTPPQFMDPG 536
DB 485 LWFSDKITDVTNMQPTDFRFPAGRTSKEAEDMTDMCVMSAKIGGLPGLSLPQFMEPG 544
QY 537 LALHAGTTRIGFPAK--TTVADNNSLVWDFANLYVAGNGTIRTFGENPTLTSCHAIK 594
DB 545 LVLLHGTHRMGFDEQEDKCCVNTDSRVFGKFLGCGGNIPATYAGANPTLTAMSLAIK 604
QY 595 SARSIIN 601
DB 605 SCEYIKN 611

RESULT 2

US-09-305-381-2
; Sequence 2, Application US/09305381
; Patent No. 6148665
; GENERAL INFORMATION:
; APPLICANT: Christensen, Soren
; APPLICANT: Lassen, Soren Flensted
; APPLICANT: Schneider, Palle
; TITLE OF INVENTION: Nucleic Acids Encoding Polypeptides
; FILE REFERENCE: 5571.200-US
; CURRENT APPLICATION NUMBER: US/09/305.381
; CURRENT FILING DATE: 1999-05-05
; EARLIER APPLICATION NUMBER: 60/088.724
; EARLIER FILING DATE: 1998-06-10
; EARLIER APPLICATION NUMBER: PA 1998 00774
; EARLIER FILING DATE: 1998-06-08
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 622
; TYPE: PRT

; ORGANISM: Coriulus Versicolor
us-09-305-381-2

Query Match 36.8%; Score 1210; DB 4; Length 622;
Best Local Similarity 44.6%; Pred. No. 2.1e-116;
Matches 274; Conservative 71; Mismatches 197; Indels 72; Gaps 16;

QY 37 DVFTAGSGPIGATYAKLCVEAGLRVVMVEIGAADSFYAVNAEEGTAVPYVPGYHKKNEIE 96
DB 48 DVFTAGSGPICTYARELVVVEAGLVVVEIGAADSFYAVNAEEGTAVPYVPGYHKKNEIE 95
QY 97 FOKDIDRFVNVKIGALQOVSVVVRNQVPTLDPGAWSPGSSAISNKNPHQREFENLS 156
DB 96 YQKNIDRFVNVQIGLMSVSVVNTLVITLSPASWQA--STFFVRNGANPEQDPLRLNS 153
QY 157 AEATRVGGGSMTHWTCTPRIHPMESLPGIGRPKL--SNDPAEDDKENNELYSEAEHLI 215
DB 154 GOAVTRVVGGMSTHWTCTATPRFEK-----LQRPVLVKNDSKADDAEMDRLYKAESYF 206
QY 216 GTSKFEDESIRHTLVLSLQADAYKDRQIRFRPLPLACHRLKNAPEYVEVHSAENLF--H 273
DB 207 KTGTQDQKESIRHNLVNLKLAEEYKG--VRDQOIPLAATR--QSPTFEVWSSANTVFDLE 263
QY 274 SIYNDKOKKLTLLTNHRCRLALTGGYKKGAAEVRNLLATRNPSQLDSYIMAKVY 333
DB 264 NRPNKDPKQFENLFPVACTNVRNDRNANSEIVG--LDVRDLHGKSKIT-----IKAKVY 316
QY 334 VLASCAIGNPQILYNSGSGGLQVTPRND-----SLIPNLGRYITEOPMAFCOIVLRQEF 387
DB 317 ILTAGAVHNAQLLAASGFGQL---GRDPAPAKPLSLPLGTHITHTQTLVFCQVMSTEL 373
QY 388 VDSVRDDP--YGLP-----WKKEAVAQIAHAKNPTDALPIPFDRPE 425
DB 374 INSVTADMTIVGKPGHPDYSVYTPGNPNKHPDKWNEKVKNNMQHQPDLPIPFDEPE 433
QY 426 PQTTPPTTEHPHWTQIHRDAFSYGAVGPEVDSRVIVDLRFWFGATDEANNLLVFQNDVQ 485
DB 434 PQTTLTFOATHPHWTQIHRDAFSYGAVQOQSDSLRILVDMFRFGTEPKEENKLFSDKIT 493
QY 486 DQSNMPOQTFYRPSASNVARKKMMADMCEVASNLGGYLPSTPPQFMDPGALHLAGTT 545
DB 494 DAYNLRQPTDFR--FPGREADEMTDMCVMSAKIGGLPGLSGYPOFMEPGLVHLGTH 551
QY 546 RIGEDKAT--TVADNNSLVWDFANLYVAGNGTIRTFGENPTLTSCHAIKARSIIINTL 603
DB 552 RMGFDEKADKCCVNTDSRVFGKFLGCGGNIPATYAGANPTLTAMSLAIKSEYI---K 608
QY 604 KGGTDGKNTGEHRN 617
DB 609 KNFEPSNPVKHN 622

RESULT 3

US-09-023-731-1
; Sequence 1, Application US/09023731
; Patent No. 6291648
; GENERAL INFORMATION:
; APPLICANT: Kawamura, Yukio; Morita,
; APPLICANT: Akihiko; Izumo, Koji.; Saka, Tomohide.
; TITLE OF INVENTION: ANTIMUR PROTEIN AND
; TITLE OF INVENTION: GENE ENCODING SAME
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORGAN & FINNEGAN, L.L.P.
; STREET: 345 PARK AVENUE
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: USA
; ZIP: 10154
; COMPUTER READABLE FORM:
; MEDIUM TYPE: FLOPPY DISK
; COMPUTER: IBM PC COMPATIBLE

Db 109 VER-----EDGVSAPWPSYAELEPYAEERLFGV-RGAGDDPTEPPRSAPYMH 158
QY 209 SEA--ERLIGTSKEFDESIRHTLVLSRQDAYKDRQIRPLPLA-----CHRLKN 258
Db 159 APIPEPVIGRAVGPER-----LGLRPFH-----MPSAIDYGPGLCRRCGT 201
QY 259 APYVWHSANLFSHIYNDKOKKLF-----TLTNHRCRTRALTGGEYKKGIAA 309
Db 202 CDAFVCRFAKG-----DAETRLRPAIRHPNVSLLTGARVRL-IADGDKHIVAV 252
QY 310 EVANLLATRNPPSOLDSDYIMAKVYVLASGAIGNPQILYNSGFSGLQVTPRNDLSLIPN--- 366
Db 253 EIER-----AGEIIT-IEAPLFVLSAGAINSALILLRSA-----DEKKNGLA 294
QY 367 -----LGRVITEOPMA-----FCQIVLRQEFVDSVRDDP----- 395
Db 295 NSSGVYGRVILMNHLSGLMGLLPFTINDTRFKTMSLNDPFGDTPGDEAAAGNVQMLGNI 354
QY 396 -----YGLPWKEAVAQHTAKNPTDALPIPRDPEQVTPPTTEHPWHTQIHRDAFSY 449
Db 355 QGPMIRAAIWPMPRPLANLARSVDVFLV-----SED----- 387
QY 450 GAYGPEVDSRVIVDLRFWFGATDPEANLLVFQNDVODGYSNPOPTFRYRSTASN-----V 505
Db 388 ---TPYDSRV---KPGKNGAE-----LIYRPGDREAHQRFV 419
QY 506 RAKMADMCVEASNLGGYIPTSPFMDPGLALHAGTTRIGDKATTVADNNSLWDF 565
Db 420 RHRSLRLKNGFPVVLGSHSGIEAPS-----HOCGTVMGDDPKAEALNALCQTYDH 471
QY 566 ANLYVAGNGTIRTFGENPTLTSCHAIK 594
Db 472 PNLYVVDAGFFPSSAALNPALTPVAAQALR 500

RESULT 14
JC6564
cellobiose oxidase (EC 1.1.3.25) precursor - white-rot fungus (Trametes versicolor)
N:Alternate names: cellobiose dehydrogenase
C:Species: Trametes versicolor (white-rot fungus)
C:Date: 05-Dec-1998 #sequence_revision 05-Dec-1998 #text_change 19-May-2000
C:Accession: JC6564
R:Dumonceanu, T.J.; Bartholomew, K.A.; Charles, T.C.; Moukha, S.M.; Archibald, F.S.
Gene 210, 211-219, 1998
A:Title: Cloning and sequencing of a gene encoding cellobiose dehydrogenase from Trametes
A:Reference number: JC6564; MUID:98248903
A:Accession: JC6564
A:Molecule type: mRNA
A:Residues: 1-768 <DUM>
A:Cross-references: GB:AF029668
A:Experimental source: 52J, AFCC 20869
C:Comment: This is a hemoflavoenzyme that oxidizes cellobiose and reduces iron(III) and
C:Genetics: cdh
A:Introns: 17/1; 103/2; 140/1; 179/1; 226/3; 281/3; 306/1; 324/2; 534/3; 577/2; 593/3; 6
C:Keywords: chromoprotein; heme; iron; metalloprotein; oxidoreductase
F:1-19/Domain: signal sequence #status predicted <SIG>
F:20-768/Product: cellobiose dehydrogenase #status predicted <MAT>
F:80,128/Binding site: heme iron (Met, His) (axial ligands) #status predicted

Query Match 4.2%; Score 138.5; DB 2; Length 768;
Best Local Similarity 19.2%; Pred. No. 0.032;
Matches 128; Conservative 83; Mismatches 228; Indels 229; Gaps 33;

QY 37 DVFIAGSGPTGATYAKLCVAGLRVVMVEIG-----AADSFYAVNAEGTAVPY-VPG- 88
Db 231 DYIVVAGPGGLVADRLSAGKKVLLERGGPSTATGCTGYDATWAKSANLTKFDVPLG 290
QY 89 -----YHKKNIEFQKIDRFVNVVNIKALQVSVFVRNQNPTLDPGAWSPGSSATNG 144
Db 291 FEFLTDNPFWCKDNTNFFAGCLLG-----GGTSV-NG 323

QY 145 K---NPHOREFENLSAEAVTRGVGGMSTHTWCSTPRHPPMPSLPGIGRKLKSLNDPAEDD 201
Db 324 ALYWPNSRDFSTAS-----GWPSWSNHOPFTDKLQRLPS-----TDHPSADG 368
QY 202 KENNELYSEA--ERLIGTSKEFDESIRHTLVLSRQDAYKDRQIRF-----RP 248
Db 369 QRYLEQSATVVOQLSQGYSQI-----TINDNPDSDKHVGFSAFDFLNCQRA 417
QY 249 LPACHRLKNAPYEVWHSANLFSHIYNDKOKKLFLLTNHRCRTRALTGGEYKKGIA 308
Db 418 GPVA-----TYFETALARKNF--VYKDN--VLVTQVIRNGSTILGVRTN-DNTLGP 463
QY 309 AEVRNLLATRNPPSOLDSDYIMAKVYVLASGAIGNPQILYNSGFSG---LQVTPRNDLSLIP 365
Db 464 DGIVPL-----NPNGRV-----ILSGSGFTPRILFOSGIGPTDMLQTVOSNAQAAA 510
QY 366 NLGRYTEQPMACQIVLROEFVDSVRDDPYGLPMMKEAVAQHTAKNPTDALPIPRDPE 425
Db 511 NL-----PP-----QSEWID-----LP-----VGQSVSDNPS----- 532
QY 426 PQVTPPTTEHP-----WHTQIHRDAFSY-----GAVGPEVDSRVIVDLRW 466
Db 533 --INLVET--HPSIDAYDNWADVMSNPRPADAAQYLOQSRGVLGASPKLNF-----WRA 583
QY 467 FGATDPEANLLYFQNDVODGYSNPOPTFRYRPS-----TASN 504
Db 584 YGSGDGITR---YAQGTVRPGAASVNTSVAINASEIPTIYLSNGIQSRIGRIGVDMALN 640
QY 505 VRA-----RKMDADMCVEASNLGGYIPTSPFMDPGLALHAGTTRIGDKATTVADNNSLWDF 565
Db 641 AKALVNPWLTVNSDKTVLLQALHDVTSTMKNVPLGTLITPDNTMTLEQYVAAVDPATMCS 700
QY 539 LHLAGTTRIGDKATTVADNNSLWDFANLYVAGNGTIRTFGENPTLTSCHAIKARS 598
Db 701 NHWVGAAGKMTSSSTAVVDENAKVFENTDNLFIVDASIIPSLPICNPOGVLM-SAAEQAVS 759
QY 599 IINTLKG 606
Db 760 RILALAG 767

RESULT 15
AD3255
l-sorbose dehydrogenase (FAD) (EC 1.1.99.-) [Imported] - Brucella melitensis (strai
C:Species: Brucella melitensis
C:Date: 01-Feb-2002 #sequence_revision 01-Feb-2002 #text_change 01-Feb-2002
C:Accession: AD3255
R:DelVecchio, V.G.; Kaputal, V.; Redkar, R.J.; Patra, G.; Mujer, C.; Los, T.; Ivan
; Mazur, M.; Goltsman, E.; Selkov, E.; Elzer, P.H.; Hagius, S.; O'Callaghan, D.; L
Proc. Natl. Acad. Sci. U.S.A. 99, 443-448, 2002
A:Title: The genome sequence of the facultative intracellular pathogen Brucella mel
A:Reference number: AD3252; PMID:11756688
A:Accession: AD3255
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-494 <KUR>
A:Cross-references: GB:AE008917; PIDN:AAL51207.1; PID:g17981893; GSPDS:GN00190
A:Experimental source: strain 16M
C:Genetics:
A:Gene: BMEI0025
A:Map position: I
C:Keywords: oxidoreductase

Query Match 4.2%; Score 138; DB 2; Length 494;
Best Local Similarity 20.4%; Pred. No. 0.017;
Matches 125; Conservative 67; Mismatches 240; Indels 182; Gaps 23;

QY 37 DVFIAGSGPTGATYAKLCVAGLRVVMVEIGADSFYAVNAEGTAVPYVPGYHKKNIE 96
Db 6 DIVIIGSIGGATMAAGLAAGADILILEAGARLADRPENRDPRAI--FORGFRPKELW 63

Db 341 FGSQEQSAMDEVTRIGYTERCVKHGFN-----QVLKGGERLQLOVESVPRNSPDH 395
QY 262 -----YVEMHSAENLFHSIYNDKOKKFLTLTNHRCRTRAL-----TGGYEKK-----I 306
Db 396 YGGLCGYGCACAGKNGTDQTLVDVAVENGAVILTGKAERFVLVDNTSSNERKKRCVGV 455
QY 307 GAAEVRNLLATNPSSOLDSTYMAKVVYVLAIGAIGNPOLYNSGSGLOVTPRNDSLIPN 366
Db 456 FASSVGGKIGK-----FIIEARVTVSSAGSLTLPMLSSGLKN-----PN 497
QY 367 LGR-----YITEOPMAFCQIVLRQEFVDSVR--DDPYGLPWKKEAVAQHIANK 412
Db 498 IGRNKLHPVMTWGYFPKDESEFGKTEGGIITSVHMND-----TESCKAILLEN 550
QY 413 PIDALPIPRDPEQVTPFTTEEBHPWHTQIHRDAFSYGAVGPEVDSRVIVDLRFWFGATDP 472
Db 551 PLIG-PASVAGLSPWVS-----GPDLMKRMK-----YGRV-- 580
QY 473 EANNLLVFQNDVQDGYSM--POPTFRYRPTASNVPRARKMADMCEVAS---NLGY--- 524
Db 581 --AHLFALVRDLGSGEVAMENEVYRTTKDRENLRAGLRQALRVSAAGAVEVGTYRSD 638
QY 525 -----LPTSPQEMDPGLAL-----HLAGTRIGFDKATTVADNN 559
Db 639 GOKMCEALITKAMEEFLDEVDVAVGVGKGYWYTFSAHQMGSCRMGVTAEEGALDEN 698
QY 560 SLVWDFANLYVAGNGTIRTGFGENTLTSMCHAIRKSARSIIINTLGGT 607
Db 699 GESWEAEGFLVCDGSLPSAVGNPMITIQSTAYCISSKIVDSLQNT 746

RESULT 10
E98192
probable oxidoreductase chain CJ0415 [imported] - Agrobacterium tumefaciens (strain C58,
C:Species: Agrobacterium tumefaciens
C:Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 11-Jan-2002
C:Accession: E98192
R:Goodner, B.; Hinkley, G.; Gattung, S.; Miller, N.; Blanchard, M.; Quorillo, B.; Goldman,
A.; Liu, F.; Wollam, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Markelz, B.;
Science 294, 2323-2328, 2001
A:Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tum
A:Reference number: A97359; PMID:11743194
A:Accession: E98192
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-561 <KUR>
A:Cross-references: GB:AE007870; PIDN:AAK89063.1; PID:g15158859; GSPDB:GN00170
C:Genetics:
A:Gene: AGR_L_980
A:Map position: linear chromosome

Query Match 4.5%; Score 148.5; DB 2; Length 561;
Best Local Similarity 19.1%; Pred No. 0.0033;
Matches 135; Conservative 60; Mismatches 222; Indels 289; Gaps 29;

QY 37 DVFIAGSGPIGATYAKLCVAGLRVYVMEIGAASFYAVNAEEGTAVPYVGYHKKNEIE 96
Db 7 DAIVVSGISGGWAANKLTQKGLKVLML-----ERGRNIEHVTDY----- 46
QY 97 FOKDIDRFVNIKALQVQSVPRVQNVPTLPDGAWSAPPGSSAISNGKNPHOREFENLS 156
Db 47 -----QNA-----DKEAWDYPHNRNATQEMKAKYVPLSRDYL 78
QY 157 AEAVTRG-----VCGMSTHTCTSPRIHPHPPMESLPGIGRPKL 193
Db 79 LEATLGMWADQETPYVEEKRFDFRGRVHVGGRSLLMGRQTYRW--SOTDFEANAOKGI 136
QY 194 SND---PAEDDKENNELYSEARLICTSTKEPDESIRHTLVLSLQDQYDKQRIFRPLP 250
Db 137 AVDWPIRVEDVSPW---DYVERFAGISG-----REGLDILPDGE-FLPPIP 180

QY 251 LACHRLKNAPEVVEVHSAENLFHSIYNDKOKKFL-----TLLTNHRCRTRALT----- 299
Db 181 L-----NEVEDQVASRL-----KKAFAKGRHLINRSRANITQELPDQERTP 221
QY 300 -----GGY-----EKKIGAAEVRNLLATR-----N 319
Db 222 CQFRNKRCLGCGPFGYFTQASTLPAAVATGNLTJLPFFSIVREILYDKKKKARGVEIID 281
QY 320 PSSOLDSTYMAKVVYVLAIGAIGNPOLYNSGF-----SGLOVTPRNDSLIPNIGRYITE-- 373
Db 282 AETNLTYEYADIIIFLNASLTNSTVLMNSATDVWEGGL-----CSSSGELGHNVMDDH 335
QY 374 -----OPMAFCQIVLRQEFVDSVR--DDP-----YGLP----- 399
Db 336 FRMGATQVGDGFEDEFYFKRRPAGE-----YIPFRFNTGDKRKYLRGFGYOGSASR 387
QY 400 --WKEAVAQHTAKNPDTALPIPRDPEQVTPPTTEEBHPWHTQIHRDAFSYGAVGPEVD 457
Db 388 SRWEIEALNIGADYKOTL-----TEPGGW--TIGMTAF--GEMLPYHD 428
QY 458 SRVIVDLRFWFGATDPEANLLVFQNDVQDGYSMPOPTFRYRPTASNVPRARKMADMCE- 516
Db 429 NRVKLD-----HDKKDKWGL-----PVLSSNVEMKQNELDMRED 462
QY 517 -----VASNLGGYILPTSPQFMDPGLALHLAGTRIGFDKATTVADNNS 560
Db 463 MYNDVAVEMPEAVGINKVPKSRGYA-----PCMGIEHMGRTARMGRDPKTSVLNGNN 513
QY 561 LYWDFANLYVAGNGTIRTGFGENTLTSMCHAIRKSARSIIINTLKG 606
Db 514 QVWDAPNFVFTDGCAMTSASCYNVSLTYMALTAARAEFAVSEKKG 559

RESULT 11
AE3094
oxidoreductase Atu4377 [imported] - Agrobacterium tumefaciens (strain C58, Dupont)
C:Species: Agrobacterium tumefaciens
C:Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 11-Jan-2002
C:Accession: AE3094
R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Wc
erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutyavin, T.; Levy, R.; Li, M.; McC
; Karp, P.; Romero, P.; Zhang, S.
Science 294, 2317-2323, 2001
A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Ka
ster, E.W.
A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
A:Reference number: AB2577; PMID:11743193
A:Accession: AE3094
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-561 <KUR>
A:Cross-references: GB:AE008689; PIDN:AAAL5171.1; PID:g17742847; GSPDB:GN00187
C:Genetics:
A:Gene: Atu4377
A:Map position: linear chromosome

Query Match 4.5%; Score 148.5; DB 2; Length 561;
Best Local Similarity 19.1%; Pred No. 0.0033;
Matches 135; Conservative 60; Mismatches 222; Indels 289; Gaps 29;

QY 37 DVFIAGSGPIGATYAKLCVAGLRVYVMEIGAASFYAVNAEEGTAVPYVGYHKKNEIE 96
Db 7 DAIVVSGISGGWAANKLTQKGLKVLML-----ERGRNIEHVTDY----- 46
QY 97 FOKDIDRFVNIKALQVQSVPRVQNVPTLPDGAWSAPPGSSAISNGKNPHOREFENLS 156
Db 47 -----QNA-----DKEAWDYPHNRNATQEMKAKYVPLSRDYL 78
QY 157 AEAVTRG-----VCGMSTHTCTSPRIHPHPPMESLPGIGRPKL 193
Db 79 LEATLGMWADQETPYVEEKRFDFRGRVHVGGRSLLMGRQTYRW--SOTDFEANAOKGI 136

Db 203 RKAEEELGLHPYDLPAANASGPTWNPYGVQMGPCNFCGFCGAGYACTWYSKASPNL--NIL 260
QY 277 NDDKQKLLFTLLNHRCTRRLALGGYEKKIGA-----AEVRNLLATRNPSOLDYSYIMAKV 332
Db 261 PALRQPTPLFELRANCNVLYKVNLDSDGROATGVTVYDAQGREIVQP-----AKL 308
QY 333 YVLASAGNAPQILYNSGSLQVTPRNDSLINLGRYITEQPMACQIVLRQEFVDSVR 392
Db 309 VIISAFQFHNVRLLLSGI--GKPYDPTGEGV--VGKNFAIQNMA-----TIKAFQKDVH 361
QY 393 DDYF-----GLPWKEAQAIAKNAKPTDALPIPRDPE 425
Db 362 TNPFTVGTGGGVAVDNADNFDHGLPFGVGGSPMW-----VNGAGSKPICGLAVPQPTPS 417
QY 426 PQVTTPTTEHPWHT---QIHRDAFSYGVAGPEVDSRVIVDLRWFGATDPEANNLLVFON 482
Db 418 -----WGSQKQAVKDAYTH-----TVSMDAHSNMTRYDNLDP 454
QY 483 DVODGYSMP--QPTFRY-----RPSTASNVRAKMMADMCEVA-----SNLGGYLPSP 530
Db 455 TYKDATGQPLRLMTFDKNEIRMSRYVTEHMKRIAEAMNPKAISVSVKRFGDHFNTRYV 514
QY 531 QFMDPGLALHLAGTTRIGEDKATTVADNNSLVWDFANLYVAGNGTIRTFGENPTLTSMC 590
Db 515 Q-----TTHLLGGAINGSDPKTSVLNRYLOSNDVHNVFVNGASAPQGTGYNPTGLVAA 568
QY 591 HAIKSARSI 599
Db 569 LAYWSAKAI 577

RESULT 6
A84260
hypothetical protein Vng1035c [Imported] - Halobacterium sp. NRC-1
C:Species: Halobacterium sp. NRC-1
C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
C:Accession: A84260
R;Ng, W.V.; Kennedy, S.P.; Mahairas, G.G.; Berquist, B.; Pan, M.; Shukla, H.D.; Lasky, S.;
Leithausen, B.; Keller, K.; Cruz, R.; Danson, M.J.; Hough, D.W.; Maddocks, D.G.; Jabil
Jung, K.H.; Alam, M.; Freitas, T.
Proc. Natl. Acad. Sci. U.S.A. 97, 12176-12181, 2000
A:Authors: Hou, S.; Daniels, C.J.; Dennis, P.P.; Omer, A.D.; Ebbardt, H.; Lowe, T.M.; Li
A:Title: Genome sequence of Halobacterium species NRC-1.
A:Reference number: A84260
A:Accession: A84260
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-529 <STO>
A:Cross-references: GB:AE004437; NID:g10580587; PIDN:AAG19445.1; GSPDB:GN00138
C:Genetics:
A:Gene: VNG1035C

Query Match 5.5%; Score 179.5; DB 2; Length 529;
Best Local Similarity 22.1%; Pred. No. 1.3e-05;
Matches 143; Conservative 53; Mismatches 236; Indels 215; Gaps 28;
QY 37 DVFIAGSGPTGATYAKLCVEAGLRVYVVEIGAADSFSYVAVNAEGTAVPYVPGYHKKNEIE 96
Db 10 DVCVIGAGAGGLIADRLASDGHVVVLEAG--PRDAADR----- 50
QY 97 FOKIDRFVNIKALGOQVSVVRNQNPTLPDPAAS-APPGSAISNGKNPHOREFENL 155
Db 51 -----RMERSIRPAHGASV-----WEMGPRDAYASTGD-----RYPL 85
QY 156 SBAVTRGVGGNSHTWCTSTPRTHPP---MESLPGIGRPLKSLNDPAEDDKENNELYSEAE 212
Db 86 NAARV-KGVGGSTLWQGMVRLHQEQDFLASATGVA-----DWPPDYDTLPKYAAAE 139
QY 213 RLIT---GTSTKEPDESIRHTLVLSRLQDAYKDRQIRFRLPLACHRLKNAPEYVHSAE 269
Db 140 SALGVSGADNFPAPPREQHPQAPPPPSYSD--SLFAD---ACESLGIAI----- 195

QY 270 NLPHSYNDKQKKLLFTLLNHRCTRRLALFG-----GYE-----K 304
Db 186 ---HSPFN-----ARLSAGRETRACVGYGTCQPCVCPGAKYDATVHVDRATDAGAR 234
QY 305 KIGAAEVRNL-----LATNPSSQLDSYIMAKVYVVLASCAIGNPOIL----- 346
Db 235 VIDEAPVORLEHDAAGDRVTGAVYATPDGTHRQS---ATEFLAAGGIETPRLLLSLS 291
QY 347 --VNSGF--SGLQVTPRNDSLIPNLRGYITEOPMAFCQIVLRQEFVDSVRDDPYGLPMW 401
Db 292 DRYPDGLANSSGL-----VGRY-----FMD----- 311
QY 402 KEVAQHIANKPTDALPIPRDPEPQVTTPTTEHPWHTQIHRDAFSYGV-----G 433
Db 312 -----HLFAGAGTLDPTTRONHVGEHT--TESHOY---DRPDGSRGAIKLEFLNYAG 360
QY 454 PEVDSRVIVDLRFGA-----TDPEANLLVFQNDQVQDYSMPQ 493
Db 361 PSFAEMALSDDMDAMCDIRDASTHIAVGLVEQPPRPNRVLHPERTDVHGNPVP 420
QY 494 TFRYRPTASNVRAKMMADMCEVASNLGGYLP-TSPQPMDFGLALHLAGTTRIGFDKA 532
Db 421 DVVWSLSAYERRTIERANEIQRBEILTELGADIEWTVGPE--DTGPAFHHMGTTRMGTDPA 478
QY 553 TVVADNNSLVWDFANLYVAGNGTIRTFGENPTLTSMCHAIKSARSI 599
Db 479 ESYVDPRLRTHDLSNLSVASSVFPPTAGAMNPTLTIAALAKAADHI 525

RESULT 7
B75608
GMC oxidoreductase - Deinococcus radiodurans (strain R1)
C:Species: Deinococcus radiodurans
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 17-Mar-2000
C:Accession: B75608
R;White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.
S.; Smith, H.O.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.
Science 286, 1571-1577, 1999
A:Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1
A:Reference number: A75250; MUID:20036896
A:Accession: B75608
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-722 <WHI>
A:Cross-references: GB:AE001862; GB:AE001862; NID:g6460468; PIDN:AAF12230.1; PID:g6.
A:Experimental source: strain R1
C:Genetics:
A:Gene: DRA0127
A:Map position: 2

Query Match 5.4%; Score 176; DB 2; Length 722;
Best Local Similarity 20.9%; Pred. No. 3.8e-05;
Matches 139; Conservative 77; Mismatches 239; Indels 210; Gaps 33;
QY 14 RSMQINGQIPKNAIHTYTGNDG----VDVFIAGSGPTGATYAKLCVEAGLRVYVVEIGA- 68
Db 186 RNSAITPYVPQ-----DGEVLEADAVVVGSGGGVIAARLAQAGKRVVLEAGGH 236
QY 69 --ADSFYAVNAEGTAVPYVPGYHKKNEIEFQKIDIRFVNIKALGOQVSVVRNQNPT 126
Db 237 YBEAHFGRELAAYQTLTYRGVYH-----PTADGNV-T 268
QY 127 LDPAWSAPPSSAISNGKNPHOREFENLSAEAVTRGVGGNSHTWCTSTPRTHPPMESLP 186
Db 269 LVAGA-----NLG-----GGSTVYVNSV-----PPRDDI- 293
QY 187 GIGRPLKSN-----DPAEDDKENNELYSEAEIRLTGTSTKEPDESIRHTLVLSRLQDA 238
Db 294 ---RQWASEHGLSDVADPGYD---RHIDAVLERM-GVSEQCSDHNGPHQ---RLVEGA 342
QY 239 YKQRQIRFRLPLACHRLKNAPEYVHSAENLFHSIYNDND---KOKKLEF----- 286

QY 393 DDPYG--LPWKE-----AVAQIAKNPTDALPIPRDPQVPTTPTTEHPWHQI 443
Db 362 DDFGDNFDHSEKELHGMAMISVQLGTRPIQASAPLPAGAP-----TWGAEP- 408
QY 444 RDAFSYG-----AVGPEVDSRVIVDLRFGATDPANLLVFNQDVQDYSMPQPTFRYR 498
Db 409 KKALYNFTRAITVGGQ-----GASLPKNNYLSLDPYKDAFGMPLRLRTYN 456
QY 499 PSTASNVRARKMAD-MCEVASNLGLYPTSPQFMD-----PGLALHLAGTRIGFDKA 552
Db 457 -FTDODRALHFIPTKTEAVARKMGVSIKAGALKDYVVPYQSTHNTGTTWGADRE 515
QY 553 TVVADNNSLVWDFANLYVAGNGTIRTFGENPTLTSCHAIKARSIIINTLKG 606
Db 516 TSVNTYLQHWADNLVYVAGNFQNSGYNPTDVGALAYRCAEGLKYHKS 569

RESULT 2
glucanate 2-dehydrogenase (FAD) (EC 1.1.99.3) catalytic chain precursor (validated) - EX
C:Species: Erwinia cyrripedii
C:Date: 05-May-2000 #sequence_revision 05-May-2000 #text_change 08-Dec-2000
C:Accession: B38575; T52554
R:Yun, D.Y.; Lee, Y.P.; Fan, J.G.
J. Bacteriol. 179, 6566-6572, 1997
A:Title: Cloning and expression of a gene cluster encoding three subunits of membrane-bd
A:Reference number: A38575; MUID:98012950
A:Accession: B38575
A:Molecule type: DNA
A:Residues: 1-615 <YUN>
A:Cross-references: GB:097665; NID:g2584859; PIDN:AAC45885.1; PID:g2584862
A:Experimental source: strain ATCC 29267
A:Note: part of this sequence, including the amino end of the mature protein, was determ
A:Accession: T52554
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-615 <YU>
A:Cross-references: EMBL:U97665; PIDN:AAC45885.1
A:Experimental source: ATCC29267
C:Genetics:
A:Start codon: GTG
C:Complex: heterotrimer, consisting out of cytochrome c chain (T52555), catalytic chain
C:Function:
A:Description: EC 1.1.99.3 [validated, MUID:98012950]
C:Keywords: FAD; oxidoreductase
F:1-22/Domain: signal sequence #status predicted <SIG>
P:23-615/Product: gluconate 2-dehydrogenase catalytic chain #status experimental <MAT>

Query Match 5.8%; Score 191.5; DB 2; Length 615;
Best Local Similarity 21.1%; Pred. No. 1.9e-06;
Matches 134; Conservative 83; Mismatches 256; Indels 163; Gaps 32;

QY 36 VDVFITAGSPIGATYAKLCVVEAGLRVWVEIGA-----ADSEYAVNAEECTAVPYPGYH 90
Db 29 DVAVVFGVGAGAIWAKELTEAGLVNVALERGHRTYDPGAYPOSIDELT-----YH 81
QY 91 KKEFEFKDIDRFVNVTKGALQVSVVRNQNVPVTLDPGANSAPPGSSAISGNKPNQR 150
Db 82 IRKLL-FQ-DLSKSVTIRHDASQAVPR----- 109
QY 151 EFENLSAEAVTRGVGMSTHWTCTSPRIHPMESLPG-----IGRPKLSNDPAEDD--KE 203
Db 110 ---QLAAFLPGTGTGGAGLHNSGVHFRVDPVELNLRSHRYEARYGNFPEGMTIQDFGVS 166
QY 204 WNEL---YSEARLLGTSTKEFDEIRHTLVLSQDAY--KDRQIRPLP-----LAC 253
Db 167 YNELEPFDDQAEKVFGTGSAN--TIKMKIGKGGNEFYAPDRSSDF-PLPAQRRTYSA 223
QY 254 HRLKNAPEVVEWH-----SAENLFHSTYND----- 279
Db 224 QLFAQAASVGYHPYDMPSSANTSGPYTNTYGAQMPCNFCGVCGYACWYKASPNVNI 283

QY 280 -----RQKKLFTLLNHRCTRLALTGTGYEKKIGA-----AEVRNLLATRNPSO--LDSYM 329
Db 284 LPALQEPKPELNNAYVLRVNLTKDKKRGATGYVLDGQREV--QPADLVILSAFQ 340
QY 330 AKVYVILASGAINP-OILYNSGFSGLQVTPRNDSLIPNL-GRYITEQPM---AFQIVLR 384
Db 341 HNVHMLLSGIGOPYNIPITNEGVGRNFAYQNIISTLKALFDKNTTNPFGIAGGAGVAD 400
QY 385 QEFVDSVRDDPYGL-----PWKEAVAQIAKNPTDALPIPRDPQVPTTPTTEHPWH 440
Db 401 DFNADNFHGPIGVGGSPFW-----VNOAGTRPVSGLPTPKGTPN-----WGS 444
QY 441 QIHRDAFSYGAVGPEVDSRVIVDLRFGATDPANLLVFNQDVQDYSMP--QPTFRYR 498
Db 445 QWK-----AAVADTYNHHISHDAH--GAHQSYRANYLQDLPNKNYVQGLLRWTFDQ 496
QY 499 PSTASNVR-ARKMADMCVEV--ASN-----LGYLTPSPQFMDPGL--ALHLAGTRIGF 549
Db 497 D---NDIRMAOPMVGKMKRITAMNPKMIIIG--AKPGCTHFTTYYOTTHMSGGAIME 551
QY 550 DKATTVADNNSLVWDFANLYVAGNGTIRTFGENPT 585
Db 552 DKPTSANRYLQSDVPNVFVGASAFPOGLGYNPT 587

RESULT 3
H87451
oxidoreductase, GMC family CC1634 [imported] - Caulobacter crescentus
C:Species: Caulobacter crescentus
C:Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 20-Apr-2001
C:Accession: H87451
R:Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg
N. J.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; K
Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001
A:Title: Complete Genome Sequence of Caulobacter crescentus.
A:Reference number: A87249; MUID:21173698; PMID:11259647
A:Accession: H87451
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-579 <STO>
A:Cross-references: GB:AE005673; NID:gl3423038; PIDN:AAK23612.1; GSPDB:GN00148
C:Genetics:
A:Gene: CC1634

Query Match 5.8%; Score 190.5; DB 2; Length 579;
Best Local Similarity 20.2%; Pred. No. 2.1e-06;
Matches 136; Conservative 83; Mismatches 270; Indels 185; Gaps 28;

QY 18 INQIQPKNAIHETGYNDGVDFIAGSGPIGATYAKLCVVEAGLRVWVVEIGAADSFYAVNA 77
Db 4 LNGARRK---NTY-----DAIVGSGITGTAAKELTEKGLKVLV----- 42
QY 78 EGTAVPYVPGYHKKNEIEFQKIDIRFVNVIKGALQVSVVRNQNVPV--LDPGAWSAP 135
Db 43 ERGPMVRHLEDY-----PTAMLDPMQSKYP 67
QY 136 PGSSA-----ISNGKNPHORE--PENLSAEAVTRG--VG 165
Db 68 QOKLPEALNAHYKQVRRTYGTYMTQTOFFVRDDHPYTEENRFDWI-----RGYHV 121
QY 166 GMSTHWTCTSPRIHPMESLPGRPKLSND---PAEDDKENNELYSEARLIGTSTKEF 222
Db 122 GRSLLWGRQSYR-HSPID-FEANAREGIADVPIRYEDLAPW--YEHVERFIGVSGQA- 175
QY 223 DESIRH-----TLVLSQDAYKDRORI-FRPLFLACHLKNAPVEVHSA----- 268
Db 176 -EGLPHFDGHQYQPMELNCVEKAFKARSEAFPERRVITGRHTALTDPTPEQLALGRTK 234
QY 269 ---ENL-----FHSIYNDKQKKLFTLLNHRCTRLALTGTG-----YEKKIGAAEVR 312

```

302 YEKKIGAAEYRNLLATRNPSQDLSYIMAKVYVVLASGAINGNPOILYNSQFSGLOVTPRND 361
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
407 -----AGRAYGVRARRPDGTTLD--VLADAVVVAAGATETPGLRRSGLG----- 450
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
362 SLIPNLGRYITEOPMAFCQILRQEFVSDRDPYGLPWKKEAVAQHIAK--NPTDALP 418
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
451 --HPLRHNHALHPAT---MLAGLFDDV-----FAM--RGVLSAAVHEFHESDGYVL 496
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
419 IPRDPPEFQVTTPTTEEHPWHTQIHRDAFSGYAGVPEVDSRVIVDLRW-----FGA 469
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
497 IEATSTPPGM-----GSMVFPVYGAEI-----LRWLDRAFOIATFGA 533
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
470 TDPEANLLVFQNDVQGYSMPOQT-PRY--RRTASNVREAR-----KMWADMCEVAS 519
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
534 -----MWADRGVGVTRSVRGTEVYVRIAPCEIAKLEVALQAIGRLLFAAGAVEVLT 585
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
520 NLGGYLTPSPQPMDCPL-----ALHLA-----GTTRIGFQKATTVADNNLSLWMD 564
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
586 G----IGAPFMSLPELODVLRRANRPSHLHAFHPTGTGAAGADEOLCPVDATGRLRG 611
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
565 FANLYVAGNCTIRGTGFENPTLSMCHAIKARSIIINLKG 605
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
642 VEGVWVADASILPSCPVPNPOLSTIMAMALAVADQTVAKVVG 682
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

```

Search completed: September 26, 2002, 11:11:41
Job time: 258 sec

```

QY 410 AKNFTDALP--IPRPDPQVTP-----FTBEHPWHT-----QIHRDAFS-YGAVGPEVD 457
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 292 SRLPLTAVTEPMPQRIIGSGVFTPAVFLGLSLAEDWTNRGDLQAWRLCGSYIGMIRP--- 348
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 458 SRVTVDLRFW-GATDPEANLLVFQNDVDCYSMPQP--TFYRSTASNVARKWMDMC 515
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 349 -RGVGSVRPLGIWEP-----LVSEKLAPEDWISLGVLTLLGQAMFAAG--ARKVIP--- 398
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 516 EVASNLGGYLPSPQPFMDPOL-----ALHLAGTTRIGFDKATTVADNNSLVDFAN 567
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 399 SISGHEGWTNDEVDENRKLPEKATNLMTIHLFTSTCPGGEHRDACAVDSYGRVGVEN 458
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 568 LYVAGNGTIRGFGENFTLTSKCHAIKSARSIIN 601
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 459 LFVADGSGVPEAPGVNPQMTMALAFRIAEALS 492
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 13
W70T_HUMAN
ID W70T_HUMAN STANDARD; PRT; 925 AA.
AC P57737;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE 70 kDa WD-repeat tumor rejection antigen homology.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OC NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Kawabata A., Hikiuji T., Kobatake N., Inagaki H., Ikema Y., Okamoto S.,
RA Okitani R., Ota T., Suzuki Y., Obayashi M., Nishi T., Shibahara T.,
RA Tanaka T., Nakamura Y., Isogai T., Sugano S.;
RL NEDO human cDNA sequencing project.;
RT Submitted (Aug-2000) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: CONTAINS 8 WD REPEATS (TRP-ASP DOMAINS).
CC -!- SIMILARITY: BELONGS TO THE CORONIN FAMILY OF WD-REPEAT PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation
CC at the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC ENBL; AKO25674; BAB15211.1;
CC InterPro: IPR001680; WD40.
CC Pfam: PF00400; WD40; 6.
CC PRINTS; PR003320; GPROTEINRPT.
CC SMART; SM00320; WD40; 6.
CC DR PROSITE; PS00678; WD_REPEATS_1; 1.
CC DR PROSITE; PS50082; WD_REPEATS_2; 4.
CC DR PROSITE; PS50294; WD_REPEATS_REGION; 2.
KW Repeat; WD repeat.
FT REPEAT 75 115 WD 1.
FT REPEAT 124 163 WD 2.
FT REPEAT 166 205 WD 3.
FT REPEAT 209 253 WD 4.
FT REPEAT 542 582 WD 5.
FT REPEAT 592 632 WD 6.
FT REPEAT 635 674 WD 7.
FT REPEAT 728 768 WD 8.
SQ SEQUENCE 925 AA; 100574 MW; A393CE973C94FA9 CRC64;

Query Match 3.08; Score 98; DB 1; Length 925;
Best Local Similarity 20.28; Pred.No.19;
Matches 88; Conservative 139; Indels 156; Gaps 21;

QY 129 PGAWSAPPGSGAISNGKNPHOREFENISAEAVTRGV-----GGMSTHWTCSTPIRHPPME 183

```

RL Adv. Dent. Res. 10:111-118(1996).
CC -1- FUNCTION: PEPTIDES DERIVED FROM THE PARENT ENAMELIN ARE COMPONENTS
CC OF ENAMEL, A UNIQUE AND HIGHLY MINERALIZED ECTODERMAL TISSUE
CC COVERING VERTEBRATE TEETH.
CC
CC -1- TISSUE SPECIFICITY: EXPRESSED BY SECRETORY-PHASE AMELOBLASTS.
CC TO THE MOST SUPERFICIAL LAYER OF THE DEVELOPING ENAMEL MATRIX,
CC WHILE LOW-MOLECULAR-WEIGHT ENAMELINS ARE OBSERVED IN DEEPER
CC ENAMELIN. PREFERENTIAL LOCALIZATION AMONG THE CRYSTALLITES IN ROD
CC AND INTERROD ENAMEL.
CC -1- DEVELOPMENTAL STAGE: EXPRESSED FROM LATE DIFFERENTIATION TO THE
CC TRANSITION STAGE.
CC -1- PTM: PARENT ENAMELIN IS PROTEOLYTICALLY CLEAVED INTO SEVERAL
CC SMALLER POLYPEPTIDES. CLEAVAGE OF N-TERMINAL REGION OF ENAMELIN
CC OCCURS SOON AFTER SECRETION.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: U52196; AAD10837.1; -
CC GlycosultedB; O97939; -
CC Signal; Enamel; Glycoprotein; Hydroxylation; Phosphorylation.
KW SIGNAL 1 38
FT CHAIN 39 1142 ENAMELIN.
FT CHAIN 39 ? 56 KDA ENAMELIN.
FT CHAIN 39 665 89 KDA ENAMELIN.
FT CHAIN 39 ? 142 KDA ENAMELIN.
FT CHAIN 39 ? 155 KDA ENAMELIN.
FT CHAIN 174 276 32 KDA ENAMELIN.
FT CHAIN 515 665 25 KDA ENAMELIN.
FT CHAIN 670 ? 34 KDA ENAMELIN.
FT CHAIN ? ? 45 KDA ENAMELIN.
FT MOD_RES 53 53 PHOSPHORYLATION (PROBABLE).
FT MOD_RES 191 191 PHOSPHORYLATION.
FT MOD_RES 216 216 PHOSPHORYLATION.
FT MOD_RES 547 547 HYDROXYLATION.
FT CARBOHYD 245 245 N-LINKED (GLCNAC. . .).
FT CARBOHYD 252 252 N-LINKED (GLCNAC. . .).
FT CARBOHYD 264 264 N-LINKED (GLCNAC. . .).
FT CARBOHYD 291 291 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 462 462 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 929 929 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1040 1040 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CONFLICT 680 680 H -> D (IN REF. 2).
FT CONFLICT 838 840 ROH -> TTI (IN REF. 2).
SQ SEQUENCE 1142 AA; 128352 MW; 938306BC87CC5FC6 CRC64;

Query Match 3.0%; Score 100; DB 1; Length 1142;
Best Local Similarity 19.2%; Pred. No. 18;
Matches 118; Conservative 65; Mismatches 210; Indels 220; Gaps 32;

QY 76 NAEETAVYVP--GYH-----KNEIEFKDIDR-----FVNVIK 110
DB 192 NEEGGN--PYFGFGYHGFGRPPYISEEMFEQDFEKPDKPTTAPTEPSVN--- 245
QY 111 ALQOVSVPRNQVPL-DP-GAWSAPPGSSAISNGKNPHQ-----EFENLSAEAVTR 162
DB 246 -----TTVPETNSTQPNAPNPRNGDTSPTGTS--GQGNPRSNPTGQGNVAVNSGQGVPR 299
QY 163 -----GVGGMSTHTCSTPRI-----HPPWESLPGIGR-----PKLNDPAEDDKW 204
DB 300 SQSPWGPRTIHTENPNINIRGFPARRQWRPPG---PAMGHRNGFPYRQNIQIQRPRW 356
QY 205 NELYSAEERLIGTSTKEFDESIRHTLVLSQDAYKDRQIRFPLPLACHRLKNAPYVE 264
DB 357 NSTLECK-----QAVRPGYTYRIVY-----GSTARSNPNYA- 390

QY 265 WHSAENLFHSIYNDKOKKLTLLTNHRCRLALTG-----GYEKKIGAAE 310
DB 391 -GNSANLRRRPEGNK-----NPMVTNVPAPGPKHGTVDQNMENIQNPREKQVSOKE 440
QY 311 VRNLATRNPS-----SOLDSYIMAKVYVLASGAIGNPQILYNSGFSGLQVTPRNDLSLPN 366
DB 441 -RTVVPTRDPSGWRNSQDYINKSNYKL-----PO-----PEDNMLVPN 479
QY 367 LG-----RYITEQPMACQVLVRQEFVDSVRDDPYGLPWWKEA---VAQ 407
DB 480 FNSIDQRENSYPRGESKRAPNSDQGTQIIPK-----GIVLEPRPIPYESETNQPELK 534
QY 408 HIAKNP--TDALPIPPRDEP-----QVTPPTTEHPHWQIHRDAFSGVAGVPEV 456
DB 535 HSAYPQVYITEGIPSPAKEHFPAGRTWQOEISPPFKED-PGROEHLPLSHG----- 587
QY 457 DSRVIDLRVFGATDPEANLLVFQNDVQDYSMP-----OPTFRYRSTASNVRRKMA 512
DB 588 -SRVHYVYDYNPYDPRENSPYLSNTWYERDSDSPNTMGQPNPHYPMWTPDPKET----- 642
QY 513 DMCEVASNLGGYLTSPQPFMDPLALHLAGTTRIGFDKATTVADNNSLVWDFANLYVAG 572
DB 643 -----IPYNEEDPIDPTGDEHFPQSR-----WDMBELSFKE 674
QY 573 NGTIRTGFGENPT 585
DB 675 DPTVRHYEGEOYT 687

RESULT 10
MTLD_KLEPN
ID MTLD_KLEPN STANDARD; PRT; 382 AA.
AC O9XBNG;
DT 30-MAY-2000 (Rel. 39, Created).
DT 30-MAY-2000 (Rel. 39, Last sequence update).
DT 16-OCT-2001 (Rel. 40, Last annotation update).
DE Mannitol-1-phosphate 5-dehydrogenase (EC 1.1.1.17).
GN MTLD.
OS Klebsiella pneumoniae.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Klebsiella.
OX NCBI_TaxID=573;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=RAY 2026;
RA Otte S., Lengeler J.W.;
RT "The mtI genes and the mannitol-1-phosphate-dehydrogenase from
RT Klebsiella pneumoniae";
RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
CC -1- CATALYTIC ACTIVITY: D-mannitol 1-phosphate + NAD(+) -> D-fructose
CC 6-phosphate + NADH.
CC -1- SIMILARITY: BELONGS TO THE MANNITOL DEHYDROGENASES FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: AF166095; AAD45386.1; -
DR InterPro; IPR000669; Mannitol_dh.
DR Pfam; PF01232; Mannitol_dh; 1.
DR PRINTS; PRO0084; MTLDHGRNASE.
DR PROSITE; PS00974; MANNITOL_DHGENASE; 1.
KW Oxidoreductase; NAD.
FT NP_BIND 3 14 NAD (BY SIMILARITY).
SQ SEQUENCE 382 AA; 41130 MW; D9927E6C068C5344 CRC64;

Query Match 3.0%; Score 98.5; DB 1; Length 382;
Best Local Similarity 22.0%; Pred. No. 4.6;

Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.
-!- FUNCTION: REQUIRED FOR BOTH FULL TRANSCRIPTION REPRESSION AND
CC ACTIVATION OF MANY GENES INCLUDING CELL TYPE-SPECIFIC GENES (STB6,
CC TY2 AND HO), CELL DIFFERENTIATION-SPECIFIC GENES (SP013), GENES
CC THAT RESPOND TO EXTERNAL SIGNALS (PH05) AND TRK2. IT IS PROBABLY
CC INVOLVED IN THE SAME TRANSCRIPTIONAL REGULATORY FUNCTION OR
CC PATHWAY AS THE TRANSCRIPTIONAL REGULATORY PROTEIN RPD3.
CC -!- SUBUNIT: IT PROBABLY FORMS A COMPLEX WITH THE TRANSCRIPTIONAL
CC REGULATORY PROTEIN RPD3.
CC -!- SUBCELLULAR LOCATION: NUCLEAR, POSSIBLY LINKED TO CENTROMERE.
CC -!- DOMAIN: CONTAINS A PAIRED AMPHIPATHIC HELIX MOTIFS, SEPARATED BY A
CC 10 TO 30 AA SEGMENT THAT FORMS POSSIBLY A LOOP, RESULTING IN A
CC STRUCTURE SIMILAR TO THAT OF HLH AND TPR MOTIFS.
CC -!- SIMILARITY: TO S.POMBE SPAC12C2.10C.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: M36822; AAA34839.1; -
CC EMBL: 274746; CA99003.1; -
CC PIR: S12068; RGV53.
CC SGD: S0005364; SIN3.
CC InterPro: IPR003822; PAH.
CC Pfam: PF02671; PAH; 3.
CC Repeat: Transcription regulation; Repressor; Cell division;
CC KW Activator; Nuclear protein.
CC FT DOMAIN 239 1200 4 X 2 PAIRED (A,B) AMPHIPATHIC HELICES.
FT REPEAT 239 252
FT REPEAT 274 285
FT REPEAT 426 439
FT REPEAT 461 472
FT REPEAT 469 692
FT REPEAT 714 725
FT REPEAT 1152 1165
FT REPEAT 1189 1200
FT DOMAIN 480 519
FT CONFLICT 510 510 Q -> QAO (IN REF. 2).
FT SEQUENCE 1536 AA; 174838 MW; 0834726312B13878 CRC64;

Query Match 3.1%; Score 102.5; DB 1; Length 1536;
Best Local Similarity 19.3%; Pred. No. 19;
Matches 115; Conservative 83; Mismatches 192; Indels 207; Gaps 31;

QY 72 FYAVNAEETAVPVPGYHKKNEIEFQKIDRFVNV---IKGALQVSVPRNVPTLD 128
DB 985 FYDILCLADFTTHTPAYSNPDKERLKLKYPISLFFSIFSEKIEESLYSHKQNV--- 1040

QY 129 PGAWSAAPGGS---SAISNGKNPHOREFENLSA-----EAVTRGVGGMSTHWTCTPRIH 179
DB 1041 ---SESSGSDGSSATSKRPYQOEMSLDLHRSRYQKLKR-----SNDEGKVPQLS 1091

QY 180 PMWESLPG-IGRPKLSNDPAEDD-----KAW 204
DB 1092 EPPEEPNTEEBELIDEEAKNPWLTGTLNVEANSQIIQNRSIFNLANTNIYIFFRHW 1151

QY * 205 NELYSAEALIGTSTEFDESIRHTLVLSL-----QDAYKD 241
DB 1152 TWIY---ERLL---EIKOMNERTVKEINTRSTVTFAKDLDLLSSQSEMGLDFYGEDAYKQ 1206

QY 242 RQIRFPLPLACHRLNAPEYVWEHSAENLFHSIYNDKOKKLTFL-----LTNHRCTR 295
DB 1207 VLRLSR-----RLINGDLEHQPW--EBSLROAYN-NKAFKIYTTDKVTQSLVKAHT- 1255

QY 296 LALTGYEKKIGAAEVRNLLATNPSSQLDSYIMAKVYVLASGAIGNPQILYNSGFSGLQ 355
DB 1256 -LWTDAKTAELNALFYKD-----RN-----ASTSAKDQIIYR-----LQ 1289

QY 356 VTPRNDSLIPNLGR-----YITEQPMAFQOIVLRQEFVDSVRDDPYGLPMWKEVAQ 407
DB 1290 VR-SHMSNTENMPRIEDFKKRLTHVSIQYIALDLTLKEPKADEDK-----WKYVVT 1340

QY 408 HIAKNTDIALPIFRDEPOVTPFTTEHHPWHQIHRDAFSGYAGVDSRVIVDLRWF 467
DB 1341 YALPHPTTEGI-----PHEKLKIPFLE-----RLIEFGQDIDGTEVDEEF----- 1379

QY 468 GATDPEANNL-----LVFQNDVQDG-YSM--POPTFRYRPSSTASNVRAKMMADMC 515
DB 1380 ---SPEGISVSTLKIKIQTPIYQIHLHENGSDYVTRKATNKY-PTTANDNTQKGMYSQKK 1435

QY 516 EVASNLGGYLPTSPQFMD--PGLALHLACTTRIGEDK-----ATTVAONNSL 561
DB 1436 ELIS-----RFLDCAVGLRNRLDEAOKLSMQKWKENLKDSIAKTSAGNOGI 1481

RESULT 7
SCA4_RICPA STANDARD; PRT: 1010 AA.
AC 09AJ75;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Antigenic heat-stable 120 kDa protein (PS120) (120 kDa antigen)
DE (protein PS 120) (fragment).
GN SCA4 OR D.
OS Rickettsia parkeri.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Rickettsia;
OX NCBI_Taxid:35792;
[1]
RP SEQUENCE FROM N.A.
RA Sekeyova Z., Roux V., Raoult D.;
RT 'Phylogenetic analysis of Rickettsia spp. by comparing sequence of the
RT 'gene D' coding for an intracytoplasmic protein'.
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (Probable).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: AF155059; AAK30690.1; -
CC Antigen.
KW NON_TER 1 1
FT NON_TER 1010 1010
FT SEQUENCE 1010 AA; 110694 MW; 78586D8C92FF9C5B CRC64;

Query Match 3.1%; Score 101; DB 1; Length 1010;
Best Local Similarity 18.2%; Pred. No. 13;
Matches 106; Conservative 88; Mismatches 243; Indels 144; Gaps 25;

QY 136 PGSAISNGKNPHOREFENLSAEAVTRGVGGMSTHWTCTSTPR---IHPPMESLPGIGRP 191
DB 55 PMSVLSGNISPSQ-----TSDPTKAV-----RETIQPKDNLIEQLKDLAAL--- 100

QY 192 KLSNDPAEDDKENNELYSEAEERLIGT-----STKEF-DESIRHTLVLSLQD-----AYKD 241
DB 101 -TORDLAEOQRKEIEBEKEKDKTLSTFFGPNANREFIDKALENPELKKELESIAGYKN 159

QY 242 RQIRFPLPLACHRLNAPEYVWEH---SAENLFHSIYNDKOKKLTFL----- 287
DB 160 VHNFTS-----AASGYPGGKPVQVQWENHVSASDLRATVKNADGDELCTNETTVTKKPT 215

QY 288 LTNHRCRLALTGGE-----KKIGAAEVRNLLATR-----NPSSQLDSYIMAKVYVL 335
DB 216 LAKQDGTQVOISSYREIDFFPIKDKADGSMHL-SWVALKADGTPKSKRPVYFAH---Y 271

OS Aspergillus oryzae.
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
 OC Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.
 OX NCBI_TaxID=5062;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=91254744; PubMed=1368680;
 RA Hata Y., Kitamoto K., Gomi K., Kumagai C., Tamura G., Hara S.;
 RT "The glucoamylase cDNA from Aspergillus oryzae: its cloning,
 RT nucleotide sequence, and expression in Saccharomyces cerevisiae.";
 RL Agric. Biol. Chem. 55:941-949(1991).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92104497; PubMed=1761224;
 RA Hata Y., Tsuchiya K., Kitamoto K., Gomi K., Kumagai C., Tamura G.,
 RA Hara S.;
 RT "Nucleotide sequence and expression of the glucoamylase-encoding gene
 RT (glua) from Aspergillus oryzae.";
 RL Gene 108:145-150(1991).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=RIB 40;
 RA Hara S., Tamura G., Kumagai C., Gomi K., Kitamoto K., Tsuchiya K.,
 RA Hata Y.;
 RL Submitted (JAN-1996) to the EMBL/GenBank/DBJ databases.
 CC -1- CATALYTIC ACTIVITY: Hydrolysis of terminal 1,4-linked alpha-D-
 CC glucose residues successively from non-reducing ends of the chains
 CC with release of beta-D-glucose.
 CC -1- SIMILARITY: BELONGS TO FAMILY 15 OF GLYCOSYL HYDROLASES.
 CC
 CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC
 DR EMBL; D01035; BAA00841.1; -
 DR EMBL; D10698; BAA01540.1; -
 DR PIR; J01346; J01346.
 DR HSP; P04064; IGAI.
 DR InterPro; IPR002044; CBD_4.
 DR InterPro; IPR000165; Glyco_hydro_15.
 DR Pfam; PF00686; CBD_4; 1.
 DR Pfam; PF00723; Glyco_hydro_15; 1.
 DR PRINTS; PR007136; GLHYDRLASE15.
 DR PRODOM; PD001368; CBD_4; 1.
 DR PROSITE; PS00820; GLUCOAMYLASE; 1.
 KW Hydrolyase; Glycosidase; Polysaccharide degradation; Glycoprotein;
 KW Signal.
 FT SIGNAL 1 19 BY SIMILARITY.
 FT PROPEP 20 25 BY SIMILARITY.
 FT CHAIN 26 612 GLUCOAMYLASE.
 FT BINDING 146 146 SUBSTRATE (BY SIMILARITY).
 FT ACT_SITE 202 202 CATALYTIC BASE (BY SIMILARITY).
 FT ACT_SITE 205 205 GENERAL ACID CATALYST (BY SIMILARITY).
 FT ACT_SITE 206 206 INTERACT WITH SUBSTRATES (BY SIMILARITY).
 FT DISULFID 236 239 BY SIMILARITY.
 FT DISULFID 248 247 BY SIMILARITY.
 FT DISULFID 288 296 BY SIMILARITY.
 FT CARBOHYD 39 39 N-LINKED (GLCNAC...) (POTENTIAL).
 SQ SEQUENCE 612 AA; 65486 MW; CD7B23E5FA978F97 CRG64;
 Query Match 3.2%; Score 103.5; DB 1; Length 612;
 Best Local Similarity 19.0%; Pred. No. 3.9;
 Matches 125; Conservative 67; Mismatches 207; Indels 259; Gaps 31;
 QY 3 LSTEQMLRDYPRSMQINGQIPKNAIHETVYGNCDGVDFIAGSGPIGATYAKLCVEAGLRVV 62
 DB 34 LSTE-----ANFSRAILNNTGADGQSAQSGAPGVVIASPSK----- 70

QY 63 MVEIGADSFYAVNAEGTAVPVPGYHKKNEIEFQKIDRFVNVIKGALQOVSVPRNQ 122
 DB 71 ----SDPDYFTYTRDSDLVGLVMKTLVDFRGGDADLLPIIEEFIS-SQARIQGISNP- 121
 QY 123 NVPTLDPGAWSAPPSSAISNGKNPHQREFENLSAAVTRGVGGMSTHWTCTPRTHPM 182
 DB 122 -----SGALSSG----- 128
 QY 183 ESLPGIGRKLSDNDPAEDDKENNELYSEARLIGTSTKEFDESI----RHTLVLSLQDAY 239
 DB 129 ----GLGEPKFNVDFTAGMORPORDGAPALRATANISFGEHLVENSHTSIATDL---- 180
 QY 240 KDRQRIERPLPLACHRLKNAPEYVEHMSAENLFHSIYNDKQKLFLLTNHRTCLALT 299
 DB 181 -----VMPV-----VRNDSLVAQYWSQGF-DLWEEVQGTSEFTVAVSHR----ALV 223
 QY 300 GG---YEKKTG-----AAEVRNLLATRNPSQLDSYIMAKVYVLAAGAIGNPOLYN 348
 DB 224 EGSSFAKTVGSSCPYCDSPAQVRCYL---QSFMTGSGYIOAN-----EG 264
 QY 349 SGFSGIQTVPNRDNLIPNLGRYITEQPMAFQIVLRQEFVDSVRDDPYGLPMKKEAVAOH 408
 DB 265 GGRSGKDIN-----TVLGSIHFPDQATC-----DDATFQPCSRALANH 304
 QY 409 IAKNPTDALPIPE-----RDPEQVTTPTTEHPWH-----TOIHRDA 446
 DB 305 --KVVTDSFSAIYAINSGRAENQAVAVGRYPEDS-----YVNGNPWELTTLAAAEQLYDAL 358
 QY 447 FSGVAGVPEVDSRIYVD--LRWF-----GATDPEANLLYFONDVQ-----DGYSMFQ 492
 DB 359 YQMDKIG-----SLAITDVLSPFPFKALYSAAATGYASTTVYKDIYSAVKAYADGVQIV 414
 QY 493 PTFPRYPTSTASNVRAKMDMCEVASNLGGYLPSPQFMDPGLALHLAGTTRIGFDKA 552
 DB 415 QT--YAASGTS-----YKTDGSO----- 435
 QY 553 TTVADNNSLVMDPANLYVAG---NGTIRTFGEN-----PTLSMCHAIKSARSINT 602
 DB 436 TSARD---LTWSYAALLTANNRRNAVVPAPWGETAATSIPTASCTTSASGTYSVVIT 490
 RESULT 5
 AMYG_ASAPK STANDARD; PRT; 639 AA.
 AC P23176;
 DT 01-NOV-1991 (Rel. 20, Created)
 DT 01-NOV-1991 (Rel. 20, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Glucoamylase I precursor (EC 3.2.1.3) (Glucan 1,4-alpha-glucosidase)
 DE (1,4-alpha-D-glucan glucohydrolase).
 GN GAI.
 OS Aspergillus awamori (var. kawachi).
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
 OC Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.
 OX NCBI_TaxID=105351;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Hayashida S., Kuroda K., Ohta K., Kuhara S., Fukuda K., Sakaki Y.;
 RT "Molecular cloning of the glucoamylase I gene of Aspergillus awamori
 RT var. kawachi for localization of the raw-starch-affinity site.";
 RL Agric. Biol. Chem. 53:135-141(1989).
 RN [2]
 RP SEQUENCE OF 494-538.
 RA Hayashida S., Nakahara K., Kuroda K., Miyata T., Iwanaga S.;
 RT "Structure of the raw-starch-affinity site on the Aspergillus awamori
 RT var. kawachi glucoamylase I molecule.";
 CC -1- CATALYTIC ACTIVITY: Hydrolysis of terminal 1,4-linked alpha-D-
 CC glucose residues successively from non-reducing ends of the chains
 CC with release of beta-D-glucose.
 CC -1- SIMILARITY: BELONGS TO FAMILY 15 OF GLYCOSYL HYDROLASES.
 CC
 CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration

Db 429 VVDANCKYHGMNDLXYVGGSSVFCTSGQANPTTT 461

Search completed: September 26, 2002, 11:11:20
Job time: 262 sec

```
Query Match      4.8%; Score 158.5; DB 10; Length 748;
Best Local Similarity 19.4%; Pred. No. 0.00058;
Matches 126; Conservative 90; Mismatches 215; Indels 217; Gaps 28;

QY 37 DVFIAGSGPIGATYAKLCVEAGLRVNVVEIGAADSFYAVNAEECTAVPYVGVYHKKNEIE 96
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 239 DAVVSGSGGGVAAANLAKAGLVLEKG---NYTADHYSGLVPSMLEYKGG--- 292

QY 97 FQKIDRFVNVIGALQOQVSVVVRNQNVPTLDPGAWSPGSSAISNGKPHQREFNLS 156
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 293 -----GLTTV----- 306

QY 157 AEAVTRGVGMSHTWCTSPRIHPMESLPGICRKLSDPAEDDKWENLYSEAEFLIG 216
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 307 GSAVG---GCTAVNWSAS---ITPDHVL-----QEW-----SEGSIKIF 340

QY 217 TSKTEF-----DE-SIRHTLVLSLODAYKQRIKRLPLACHRL-----KNAPE-- 261
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 341 FGQEQYQSAMDVETIRIGVTCVKGQFN-----QVLKRCGERLGLQVESVPRNSPEDH 395

QY 262 -----YVEHSAENLFHSIYNDKQKFLTLTNHRCRTRAL-----TGGYEKK-----I 306
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 396 YCGLGCGYGRAGAKNGTDQTLWDVAVENGAVILTIKAERFVLVDNTSSNERKKRCVG 455

QY 307 GAAEVRNLLATRNPSQSDSYIMAKVYVLASGAIGNPOILYNSGFSGLQVTPRNDSLIPN 366
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 456 FASVSGKIGKK-----FIEARVTVSSAGSLTTPMLSSGLKN-----PIN 497

QY 367 LGR-----YITEOPMAFCQIVLRQEFVDSYR---DDPYGLPWKREAVAQHIANK 412
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 498 IGRNKLHPVMTWGVFPKDESEFSGKMEGGIITSVHMND-----TESGCKAILEN 550

QY 413 PTDALIPPRDPQVTPPTTEHPHWTQIHRDAFSGAVGPEVDSRVIVDLRWFQATDP 472
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 551 PLIG-PASYAGLSPPWS-----GPDLKEMIK-----YGRY-- 580

QY 473 EANNLVFQNDVDGYSM---POPTFRYRPTASNVRARKWADMCEVAS---NLGGY--- 524
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 581 --AHLFALYRDLGSGEVMEYVTRTKDRENLRAGLRQALRVSVAAGAVEGVYTRSD 638

QY 525 -----LPTSPPOFMDPGLAL-----HLAGTTRIGFDKATTVADN 559
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 639 GQKMKCEAITKEAMEEFLDEVDVAGVGVTGKEYWTVYFSAHOMGSCRMGVTAEBGALDEN 698

QY 560 SLYVDNFANLVACNGTIRTCFGENPILTSCHAIKSARSILNTLKGT 607
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 699 GESWEAEGLFVCDGSLPSAVGNPMITIOSTAYCISKIVDSLQNK 746

RESULT 13
Q988P2 PRELIMINARY; PRT; 499 AA.
AC Q988P2;
DT 01-OCT-2001 (TrEMBLrel. 18, Created)
DT 01-OCT-2001 (TrEMBLrel. 18, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE MLR6655 PROTEIN.
GN MLR6655.
OS Rhizobium loti (Mesorhizobium loti).
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Phyllobacteriaceae; Mesorhizobium.
OX NCBI_TaxID=381;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NAFF303099;
RX MEDLINE=21082930; PubMed=11214968;
RA Kaneko T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,
RA Watanabe A., Idesawa K., Ishikawa A., Kawashima K., Kimura T.,
RA Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsuno A.,
RA Mochizuki Y., Nakayama S., Nakazaki N., Shimpo S., Sugimoto M.,
RA Takeuchi C., Yamada M., Tabata S.;
RT *Complete genome structure of the nitrogen-fixing symbiotic bacterium
```

```
RT Mesorhizobium loti.
RL DNA Res. 7:331-338(2000).
DR EMBL; AP003009; BAB52905.1; -.
DR InterPro; IPR00205; NAD_Binding.
KW Complete proteome.
SQ SEQUENCE 499 AA; 54537 MW; 633924AED96B8F51 CRC64;

Query Match      4.7%; Score 155.5; DB 16; Length 499;
Best Local Similarity 20.7%; Pred. No. 0.00053;
Matches 129; Conservative 65; Mismatches 235; Indels 195; Gaps 27;

QY 37 DVFIAGSGPIGATYAKLCVEAGLRVNVVEIGAADSFYAVNAEECTAVPYVGVYHKKNEIE 96
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 7 DIVIGSGIGSSLAADTGRIVILE-----RGEHLRTP 45

QY 97 FQKIDRFVNVIGALQOQVSVVVRNQNVPTLDPGAWSPGSSAISNGKPHQREFNLS 145
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 46 ARDDIAIFQNGFYRSSE-----WLATGSEFLPGNYVYVVGNSKFFG 88

QY 146 -----NPHQREFENLSAEAVTRGVGGMSTHTWCTSPRIHP---PMSLPGICRP 191
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 89 AMYRYRQEDFNPRD-----MGRSPGWPISYAELEPWYERAILFGV-RG 134

QY 192 KLSNDPAEDDKWENLYSEAEFLIGTSTKEFDESIRHTLVLSLODAYKQRIKRLPL 251
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 135 DARQDPTPEPR-NRPY-----RYLPVDEPAIATVRRORLQAGIHPA-----SLPL 179

QY 252 A-----CHRLKNAPEYVEHSAENLFHSIYNDKQKFLTLTNHRCRTRALTTGGYEKK 305
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 180 AIDIDAMLRRAKTG-----WDAPNT--GAGKIDAEVGPLTKALEHPNATLITGANVORL 232

QY 306 IGAVERNLLATRNPSQSDSYIMAKVYVLASGAIGNPOILYNSGFSGLQVTPRNDSLIP 365
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 233 VTDASGRVMAAVFIKDGVELSIGADVFAVAGAVQAALLRS-----STSVYP 282

QY 366 N-----LGR-YITEQPMACQI-----VLQRQ-----FVDSVRDPPYG-LPWKREAV 405
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 283 NGLGSSDOLGRNEMNHTTAMLAIDPFRRNTAVYQKTLGFNDFYKDPGLSGPLGNVOL 342

QY 406 AQHIKAPTD-----LPIPRDPQVTPPTTEHPHWT-QIHRDAFSGAVG---PEVD 457
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 343 LGHITGNILKANAPLP-----RWLAGLVARNCYGFWLTSDELNP 384

QY 458 SRVIVDLRWFQATDPPEANLLVFQNDVDGYSMPOPTFRYRPTASNVRARKM--ADM 515
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 385 SRVTI-----RNGRIYVNVVRNMGAHETLIR-----RTRAVNREAGFP 423

QY 516 EVASNLGGYLPSPQPMDFGLALHLAGTTRIGFDKATTVADNNSLVDFANLYVAGNGT 575
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 424 VLTRTFGRKTT-----HQCGRARLSDPNTSVVSPDCRSHDIANLYVTDAV 472

QY 576 IRTGFGENPILTSCHAIKSARS 599
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 473 LPTSAAVNPALTITAAALAKAGAAI 496

RESULT 14
O74253 PRELIMINARY; PRT; 769 AA.
AC O74253;
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE CELLOBIOSE DEHYDROGENASE (EC 1.1.99.18).
GN CDH.
OS Pycnoporus cinnabarinus.
OC Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Homobasidiomycetes;
OC Stereales; Schizophyllaceae; Pycnoporus.
OX NCBI_TaxID=5643;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99321795; PubMed=10393235;
```

QY 556 ADNNLSLVDFANLYVAGNGTIRGTGFGNPTLTSMCHAIKARSINTLK 604
 Db 502 VDGNCRTFDEHNLWLPGGGAIPSAVSVNSTLSMAALGLKRAHDISLRMK 550

RESULT 9
 Q9HOR8 PRELIMINARY; PRT; 529 AA.
 AC Q9HOR8
 DT 01-MAR-2001 (Tremblrel. 16, Created)
 DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
 DE 01-DEC-2001 (Tremblrel. 19, Last annotation update)
 DE VNG1035C.
 GN VNG1035C.
 OS Halobacterium sp. (strain NRC-1).
 OC Archaea; Euryarchaeota; Halobacteriales; Halobacteriaceae;
 OC Halobacterium.
 OX NCBI_TaxID=64091;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=20504483; PubMed=11016950;
 RA Ng W.V., Kennedy S.P., Mahairas G.G., Berquist B., Pan M.,
 RA Shukla H.D., Lasky S.R., Baliga N.S., Thorsson V., Sbrogna J.,
 RA Swartzell S., Weir D., Hall J., Dahl T.A., Welti R., Goo Y.A.,
 RA Leithauser B., Keller K., Cruz R., Danson M.J., Hough D.W.,
 RA Maddocks D.G., Jablonski P.E., Krebs M.P., Angevine C.M., Dale H.,
 RA Isenbarger T.A., Peck R.F., Pohlschroder M., Spudich J.L., Jung K.-H.,
 RA Alam M., Freitas T., Hou S., Daniels C.J., Dennis P.P., Omer A.D.,
 RA Ebhardt H., Lowe T.M., Liang P., Riley M., Hood L., Dassarma S.;
 RT "Genome sequence of Halobacterium species NRC-1";
 RL Proc. Natl. Acad. Sci. U.S.A. 97:12176-12181(2000).
 DR EMBL; AB005037; AAG19445.1;
 DR InterPro: IPR000425; MIP.
 DR InterPro: IPR000205; NAD_binding.
 DR PROSITE; PS00221; MIP; UNKNOWN_1.
 KW Complete proteome.
 SQ SEQUENCE 529 AA; 56633 MW; FAED92051336225D CRC64;

Query Match 5.5%; Score 179.5; DB 17; Length 529;
 Best Local Similarity 22.1%; Pred. No. 6.6e-06;
 Matches 143; Conservative 53; Mismatches 236; Indels 215; Gaps 28;

QY 37 DVFIAGSGPGTATYKLCVAGLRVWVEIGAADSFYVNAEGTAVPYVPGYHKKNEIE 96
 Db 10 DVCVIGAGPAGGLIADRLASDGHSHVVLEAG--PRFDAADR----- 50

QY 97 FQDIDRFVNVIGALQOVSVYVNVQNVPLDPCAWS-APPGSSAISNGKNPHOREFENL 155
 Db 51 -----RMERSIRPAHGASV-----WEMGGPRDAYASTGD-----RYIPL 85

QY 156 SAEAVTRGVGSMTHWTCTSPTRTHPP---MESLPGIGRPKLSNDPAEDDKWENLYSEAE 212
 Db 86 NAARV-KGVGGSTLHWGQWYRLHEQDFRASATGVGA-----DWPFYDTLKYIAAAE 139

QY 213 RLI---GTSTKEPDESIRHTLVLSLDQAYKDRQIRPLPLACHRLKNAPEYVHSAE 269
 Db 140 SALGVSGASDNPAPPREOPHPQAPFPYSYSD--SLPAD--ACESLGIAT----- 185

QY 270 NLFHSYNDKQKLLFTLNHRCRTRALTG-----GYE-----K 304
 Db 186 ---HSPVN-----ARLSAGRETRACVGYGTCQVPCPGAKYDATVHVDRATDAGAR 234

QY 305 KIGAAEVRNL-----LATRNPSSQLDSYIMAKVYVLASGAIGNPQIL----- 346
 Db 235 VIDEAPVORLEHDAAGDRVTGAVATPDGTHRQS-----ATEFLAAGIETPRLLILSDS 291

QY 347 --YNSGF---SGLVTPRNDLIPNLGRYITEQPMFACQIVLROEFYDSVRDDPYGLPW 401
 Db 292 DRYPDGLANSSGL-----VGRY-----FMD----- 311

QY 402 KEVAQHIANKPTDALPIPRDPEDPQVTTPTTEHPHWHQIHRDAFSYGAV-----G 453

Db 312 -----HLFACAGGTLDEPTRQNHVGENT--TESHQYV---DRPDGSRGAIKLEFLNYAG 360
 QY 454 PEVDSRVIVDLRWEGA-----TDPEANNLLVFQNDVQDSYSNPQP 493
 Db 361 PSAPMALSGDDWDGDMCDRIIRDASGTHIAGVGLVEOQPRPENRVRLHPERTDVHGNPVP 420
 QY 494 TFRYRPTASNVARKMMADMCEVASNLGGYLP--TSPPOFMDPGLALHLAGTTRIGFDKA 552
 Db 421 DVVWSLSAYERBTIERANEIQREILTELGADEIWTGVE--DTGPAFHMGTTTRMGTDPA 478
 QY 553 TTVDNNSLVWDFANLYVAGNGTIRGTGFGNPTLTSMCHAIKARSIS 599
 Db 479 ESVYDPRLRTHDLSNLSVASSVPEPTAGAMNPTLTIAALAKAADHI 525

RESULT 10
 Q9RZ26 PRELIMINARY; PRT; 722 AA.
 AC Q9RZ26;
 DT 01-MAY-2000 (Tremblrel. 13, Created)
 DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
 DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
 DE GMC OXIDOREDUCTASE.
 GN DRA0127.
 OS Deinococcus radiodurans.
 OC Bacteria; Thermus/Deinococcus group; Deinococcales; Deinococcus.
 OX NCBI_TaxID=1299;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=RI;
 RX MEDLINE=20036896; PubMed=10567266;
 RA White O., Eisen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,
 RA Dodson R.J., Haft D.H., Gwinn M.L., Nelson W.C., Richardson D.L.,
 RA Moffat K.S., Qin H., Jiang L., Pamphile W., Crosby M., Shen M.,
 RA Vamathevan J.J., Lam P., McDonald L., Otterback T., Zaleski C.,
 RA Makarova K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,
 RA Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,
 RA Fraser C.M.;
 RT "Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1";
 RL Science 286:1571-1577(1999).
 DR EMBL; AE001862; AAF12230.1;
 DR TIGR; DRA0127;
 DR InterPro: IPR001064; Crystallin.
 DR InterPro: IPR001199; Cyt.B5.
 DR InterPro: IPR000172; GMC_oxred.
 DR InterPro: IPR001899; Gram_pos_anchor.
 DR Pfam; PF00732; GMC_oxred; 1.
 DR PROSITE; PS00225; CRYSTALLIN_BETAGAMMA; UNKNOWN_1.
 DR PROSITE; PS00191; CYTOCHROME_B5_1; UNKNOWN_1.
 DR PROSITE; PS00624; GMC_OXRED_2; 1.
 DR PROSITE; PS00343; GRAM_POS_ANCHORING; UNKNOWN_1.
 KW Complete proteome.
 SQ SEQUENCE 722 AA; 75376 MW; 450DF1CD1B7596F9 CRC64;

Query Match 5.4%; Score 176; DB 16; Length 722;
 Best Local Similarity 20.9%; Pred. No. 2.1e-05;
 Matches 139; Conservative 77; Mismatches 239; Indels 210; Gaps 33;

QY 14 RSMOINGQIPKNAIHTHYGNDG-----VDVFIAGSGPGTATYKLCVAGLRVWVEIGA- 68
 Db 186 RNSAITPYVQ-----DGEVLEADAVVGGSGGVIAARLAQAGKRVVLEAGGH 236

QY 69 --ADSFYVNAEGTAVPYVPGYHKKNEIEFQKIDRFVNVIKGALQOVSVYVNVQNVPT 126
 Db 237 YHEAHFOGRELAAYQTLYYRGGYH-----PTADGNV-T 268

QY 127 LDPGAAAPPSSAISNGKNPHOREFENLSAEAVTRGVGGMSTHWTCTPRIHPMESLP 186
 Db 269 LVAGA-----NLG-----GGSTVWNSV-----PPRDDI- 293


```
Db 154 GOATRVVGMSTHWTGATPRDREQ-----RPLLVKDDQDADDAEWDRLYTKAESYF 206
QY 216 GTSTKEFDESIRHTLVLSLQDAYKDRIPRPLPLACHRLKNAPEVVEHSAENLF--H 273
Db 207 KTGTDQFQKESIRHNLVNLKLAEEYKG-QRDQOQIPLAATR--RSPTFVWSSANTVDLQ 263
QY 274 SIYNDKQKFLTLTNRCTRRLALTGGYEKKIGAAEVRNLLATRNPSQLDSY----- 327
Db 264 NRPNTDAPNERENLFPVACERV-----VRN-----TGNSETESLRIHDLI 304
QY 328 -----IMAKVYVLASGAIGNPQILYNSGSLG-OVTPRN-DSLIPNLGRVITTEQPAFC 379
Db 305 SGRFEKADVFLTAGVHNAQLVNSGFGQLGRPPANPQLLPSLGSYITEQSILVFC 364
QY 380 QIVLRQEFVDSRDDPY--GLP-----WKKEVAOAHIAKNPTDAL 417
Db 365 QTVNSTELIDSVKSDMIIRGNPGDLGYSVTYTPCAETNKHPDMNEKVKHMMQHQEDPL 424
QY 418 PIPRDRPEQVTPPTTEHFWHTQIHRDAPSYGAVGEVDSRVIVDLRWFGATPEANNL 477
Db 425 PIPFEDPEQVTLFQPSHPWHTQIHRDAPSYGAVQOQSIDSLRIVDRFFETEPKEENK 484
QY 478 LVFQNDVQDGYMPOPTFRYR-PTASNVARKKMMADMCEVASNLGGVLPSPPOFMDPG 536
Db 485 LWFSDKITDYNMPOPTDFRFPAGRTSKEADMTDMCVMSAKIGFLPGSLFQFNEPG 544
QY 537 LAHLACTTRIGEDKA--TTVADNNSLVDFANLYVAGNGTIRTFGGENPRLTSMCHAIK 594
Db 545 LVHLAGLTHRMGEDEQDKCCVNTDSRVFGKFLNGFCGCGNPTAYGANPTLTMTSLAIK 604
QY 595 SANSIIN 601
Db 605 SCEVIKN 611

RESULT 2
Q9RH54 PRELIMINARY; PRT; 551 AA.
AC Q9RH54;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE DEHYDROGENASE SUBUNIT 1.
OS Pantoea agglomerans.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Pantoea.
OX NCBI_TaxID=549;
RN [1]
RP SEQUENCE FROM N.A.
RA Chang K.-I., Yun D.-Y., Pan J.-G., Shin Y.-C.;
RT "Cloning and expression of a gene cluster encoding three subunits of
RT membrane-bound 2-keto-D-glucanate dehydrogenase from Erwinia herbicola
RT ATCC08111 in Escherichia coli."
RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF068066; AAF21261.1;
DR InterPro; IPR000205; NAD_Binding.
SQ SEQUENCE 551 AA; 60047 MW; FE2ECCD0933DAEFB CRC64;

Query Match 6.2%; Score 203; DB 2; Length 551;
Best Local Similarity 19.9%; Pred. No. 8.7e-08;
Matches 132; Conservative 84; Mismatches 250; Indels 198; Gaps 26;

QY 24 KNAIHTYTGNDGVDFVTFAGSGPIGATYAKLCVEAGLRVWVVEIGAADSFYAVNAEETAV 83
Db 2 KKPVFTAGQDASADIVTGVSGIVGGMANELVSQGYSLVLEAGL----- 46
QY 84 PYVPVGHKKNEIFQKIDIDRFVNYKIGALQOVSVPRNQ---NVPTLDPGAWSAP----- 135
Db 47 -----RIDRAQAVENRN-----MPFANRAGSDFOGLYPOSKFAPAPLYF 86
QY 136 PGSSAISNGKNHOREFENLUSAEATRVGVGMSTHTWCSTPRTHPP---WESLPGIGRPK 192
Db 136 PGSSAISNGKNHOREFENLUSAEATRVGVGMSTHTWCSTPRTHPP---WESLPGIGRPK 192
```

```
Db 87 PRNNYV-NVTGPNADSFQ-----QGYLRTVGCTTWHAAASCHRRHPSDFVMQSYGVGR-- 139
QY 193 LSNDAEDDKEM---NEL---YSEAERLIGTSKEDESIRHTLVLSLQDAYKDRORI 245
Db 140 -----DWPYGYDELEPWYCKAENEIGVAGP--NDPARQSPTERS----- 176
QY 246 FRPLPLACHRLKNAPEVVEHSAENLFHSIYNDKQKFL-----FTLLTNH 291
Db 177 -QPYPM-----DMVPPAHGDNYPASVYVPHGYNLVPPIQGRSTRPWEGRPTCCGNN 226
QY 292 RCTRLATGGYEKKIGAAEVRNLLATRNPSQLDSYIM----- 329
Db 227 NCQPICPIGAMYNGIHVE-----RAERNGAVYLAEVVYKVMYKMDTSNNRITAVHLLDTSGA 282
QY 330 -----AKVYVLASGAIGNPQILYNSGSLGQVTPRNDSLIPN---LGRYITEQPMFACQ 380
Db 283 SHKATAKAFALACNGIETPRLLL-----MAANDANPANGIANASDMVGRNMDHSGFCS 336
QY 381 IVLROEFVDSVRDDPYGLPMWKEAVOAHIAKNPT-DALPIFRDPEQVTPPTTEHPWH 439
Db 337 FLTRK-----PVN-----LGKGAQSSCMVGYRQD-----PRRDYSAN 370
QY 440 TQI-----HRDAFSYGAVGPEVDSRV-----IVDLRWFGATDPEANNLLVFQND 483
Db 371 KVILNNSRVVTATQQAAMKGLVKGALDEEIRYRAVHSVDSLISLEPLDPDENRLTSLKT 430
QY 484 VQDGYMPOPTFRYRPTASNVARKKMMADMCEVASNLGGVLPSPPOFMDPGGLAL--HL 541
Db 431 RKDPHGLPCPDYIYDVGIVYRKGAEASHAQLHI-----GOLFDAKEFTISQGLNANNHI 485
QY 542 AGTTRIGEDKATTVADNNSLVDFANLYVAGNGTIRTFGGENPRLTSMCHAIKARSIIIN 601
Db 486 MGVIINGKNAKEAYVDCNGRAFDHENLWLPGGGAIPASVYVNSTLTMAALGLKAHDISL 545
QY 602 TLKG 605
Db 546 RMKG 549

RESULT 3
Q9PI90 PRELIMINARY; PRT; 573 AA.
AC Q9PI90;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE PUTATIVE OXIDOREDUCTASE SUBUNIT.
GN C30415.
OS Campylobacter jejuni.
OC Bacteria; Proteobacteria; epsilon subdivision; Campylobacter group;
OC Campylobacter.
OX NCBI_TaxID=197;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NCTC 11168;
RX MEDLINE=20150912; PubMed=10688204;
RA Parkhill J., Wren B.W., Mungall K., Ketley J.M., Churcher C.,
RA Basham D., Chillingworth T., Davies R.M., Feltwell T., Holtroyd S.,
RA Jagers K., Karlyshev A.V., Moule S., Pallen M.J., Penn C.W.,
RA Quail M.A., Rajandream M.A., Rutherford K.M., van Vliet A.H.M.,
RA Whitehead S., Barrell B.G.;
RT "The genome sequence of the food-borne pathogen Campylobacter jejuni
RT reveals hypervariable sequences."
RL Nature 403:665-668(2000).
DR EMBL; AL139075; CAB74251.1;
KW Complete proteome.
SQ SEQUENCE 573 AA; 63686 MW; 7BEB522A779A5F94 CRC64;

Query Match 6.1%; Score 200.5; DB 16; Length 573;
Best Local Similarity 20.8%; Pred. No. 1.5e-07;
Matches 135; Conservative 74; Mismatches 270; Indels 175; Gaps 26;
```